

**DELGAN-BC: Enhancing Breast Cancer Detection  
Through Deep Ensemble Learning and Generative  
Adversarial Networks**



**Ph.D. Thesis**

Dilawar Shah  
151-FBAS/PHDCS/F16

**Supervisor**

Dr. Mohammad Asmat Ullah Khan  
Professor, DCS, IIU

**Co-Supervisor**

Dr. Mohammad Abrar  
Assistant Professor, DCS, BKUC

**Department of Computer Science  
Faculty of Computing and Information Technology  
International Islamic University, Islamabad, Pakistan**

**2024**



Accession No. TH-27506 40

8.  
006.112  
D12

1. The following is a list of the  
names of the persons who  
were present at the meeting  
held on the 1st of June, 1964.

*A dissertation submitted to the*  
*Department of Computer Science*  
*International Islamic University, Islamabad*  
*as a partial fulfillment of the requirements*  
*for the award of the degree of*  
*Doctor of Philosophy in Computer Science.*

**INTERNATIONAL ISLAMIC UNIVERSITY ISLAMABAD  
FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY  
DEPARTMENT OF COMPUTER SCIENCE**

Date: 30 Aug, 2024


**Final Approval**

It is certified that we have read this thesis, entitled "DEIGAN-BC: Enhancing Breast Cancer Detection Through Deep Ensemble Learning and Generative Adversarial Networks" submitted by Mr. Dilawar Shah, Registration No. 151-FBAS/PHDCS/F16. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University Islamabad for the award of the degree of PhD in Computer Science.

**Committee**

**External Examiner:**

Dr. Ayyaz Hussain,  
Professor,  
Department of Computer Science,  
Quaid-i-Azam University University, Islamabad



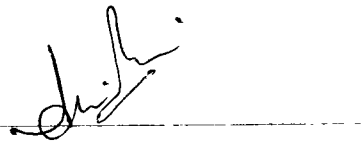
**External Examiner:**

Dr. Hassan Mujtaba,  
Professor,  
FAST University, Islamabad



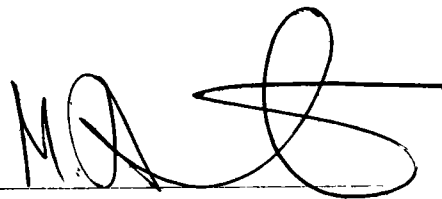
**Internal Examiner:**

Dr. Asim Munir,  
Assistant Professor,  
Department of Computer Science,  
FCTE, IUI



**Supervisor:**

Dr. Mohammad Asmat Ullah Khan,  
Professor,  
Department of Computer Science,  
FCTE, IUI



**Co-Supervisor:**

Dr. Mohammad Abrar,  
Assistant Professor,  
Department of Computer Science,  
Bacha Khan University, Charsadda



## Declaration

I, Dilawar Shah, hereby declare that my Ph.D. thesis titled “*DELGAN-BC: Enhancing Breast Cancer Detection Through Deep Ensemble Learning and Generative Adversarial Networks*” is my work, neither as a whole nor as a part thereof has been copied out from any source except where due reference is made in the text. It is further declared that I have not previously submitted the work presented in the thesis report for partial or full credit for the award of degree at this or any other university.

**Dilawar Shah**

**151-FBAS/PHDCS/F16**

## Acknowledgments

In the name of Allah, the Most Gracious, the Most Merciful, I offer my heartfelt gratitude for the unwavering strength and blessings from the Almighty Allah that have guided me through the arduous journey of completing this thesis. I appreciate my esteemed supervisor, Professor Dr. Mohammad Asmat Ullah Khan, for his consistent encouragement, invaluable technical and moral support, and constructive suggestions throughout this research. I am immensely fortunate to have had the support of precious individuals who graciously dedicated their time amidst their busy schedules to assist me in my research. I extend my whole-hearted gratitude to my co-supervisor, Dr. Mohammad Abrar, Assistant Professor Department of Computer Science at Bacha Khan University Charsadda, for his patient guidance, content validation, enthusiastic encouragement, and valuable critiques of this research work. I extend my enthusiastic thanks to the entire Department of Computer Science, Faculty of Computing & Information Technology International Islamic University, Islamabad team. I owe a deep debt of gratitude to my fellow Ph.D. colleagues, including Mr. Faizan Ullah, Mr. Sabir Shah, Mr. Shujaat Ali, Dr. Abdusalam, and Mr. Nadeem Sarwar, for their continuous encouragement and support from the inception to the culmination of this significant endeavor. Words always fall short when expressing my appreciation for the boundless love, constant support, and encouragement provided by my late father, wife, children, brothers, and sister, who have consistently stood by my side whenever I needed them. I extend my profound gratitude to Prof. Dr. Abdus Sattar, the former Dean of the Faculty of Science, Arts, and Humanities at Bacha Khan University, Charsadda, for his motivation and encouragement at the outset of this journey. Special thanks go to Pir Masoom Shah, Lecturer at DCS Bacha Khan University, Charsadda, whose valuable insights and discussions resolved many challenges throughout my PhD. I appreciate all my departmental colleagues and friends, especially Prof. Dr. Shah Khusro, for their thought-provoking ideas that significantly contributed to the success of this endeavor. I am also profoundly thankful to all clinicians and doctors for their generous allocation of time, invaluable guidance, and unwavering assistance throughout this study. Your expertise and support have played a pivotal role in completing this research.

Dilawar Shah

## Dedication

*To my late father, family, and friends, whose steadfast encouragement and support have propelled me forward in my academic career. Your support and affection for me have enabled me to do this.*

## ABSTRACT

Breast cancer is a subtype of cancer in which breast cells grow out of control. It is the second leading cause of death in women. In the last five years, there were 7.8 million women have been diagnosed with this deadly disease. Mammography is a firsthand tool for the diagnosis of breast cancer in its initial stage. It requires effective early detection methods to reduce its death toll. Healthcare specialists often face challenges in detecting small, abnormal, and concealed masses in the early stages of breast cancer. This difficulty arises from the subpar quality of mammographic images and the low sensitivity of these images in dense breast tissue. These challenges often lead to high rates of false positives and negatives, necessitating an urgent need for advanced computer-aided diagnosis (CAD) systems. It is crucial to accurately detect a breast tumor during its early stages to prevent incorrect decisions and their potentially harmful consequences. At the same time, these CAD systems need sufficient and annotated mammographic data for training. Annotating data for training is costly and time-consuming. The current CAD systems for breast cancer detection have deficiencies, such as elevated rates of false positive results and limited levels of correctness. Improving CAD systems' performance is crucial to overcoming their current automated breast cancer detection limitations.

The research focuses on overcoming data scarcity in mammographic images by employing Deep Convolutional Generative Adversarial Networks (DCGAN) for data augmentation. This approach generates realistic and diverse mammograms using the Digital Database for Screening Mammography (DDSM) benchmark dataset. Furthermore, a unique methodology for validating these synthetic images using similarity assessment and standard deviation is introduced, ensuring the integrity of the augmented dataset. The technique successfully generated high-quality synthetic mammograms for each class. The validation process assessed the similarity of each class, eliminating redundant and dissimilar images from the synthetic data. Synthetic images are further validated through expert radiologist assessments.

The study develops a Two-pathway EfficientViewNet model specifically tailored for mammograms' unique Craniocaudal (CC) and Mediolateral Oblique (MLO) views to optimize the diagnostic process. The distinctive architecture consistently enhances the performance of the EfficientViewNet model. This model enables the extraction of relevant features from each view, significantly improving the accuracy of breast tumor



classification. The EfficientViewNet model effectively optimizes its learning rate for the corresponding perspectives. The EfficientViewNet architecture achieved superior results and reduced computational complexity compared to other variants of deep neural networks. This methodology outperforms in various critical areas, with an impressive F1 score of 97.0%, precision of 98.0%, and recall value of 96.0%.

In addition, an ensemble classifier combining EfficientNet, DenseNet, ResNet, and AlexNet is devised to investigate the effects of ensemble learning on classification performance. The study uses established Radiological Society of North America (RSNA) and DDSM benchmark data sets alongside a new synthetic data set generated by DCGAN, enriching the breadth of data for model training and testing. The ensemble approach amalgamates the predictive capabilities of individual models, achieving superior accuracy and demonstrating the constructive collaboration of combined classifier wisdom. The ensemble model reported a precision of 94.6%, a sensitivity of 92.4%, a specificity of 96.1%, and an AUC of 98.0%. In individual models, EfficientNet performed well with 93.2% accuracy, 92.1% sensitivity, 93.5% specificity, and 94.0% AUC score.

Integrating Class Activation Maps (CAM) to visualize cancerous cells enhances the interpretability of the results, making them more understandable to domain experts and bridging the gap between complex AI models and clinical applications.

This thesis signifies notable progress in detecting breast cancer by employing an innovative data augmentation technique using DCGAN and optimizing Convolutional Neural Network (CNN) architectures. The research aims to improve the accuracy and reliability of mammogram classification by addressing the challenges of limited data availability and the need for model interpretability. The implementation of stringent validation protocols for synthetic mammograms and the utilization of class activation maps to enhance interpretability serve to connect intricate Artificial Intelligence (AI) models with clinical practice. These contributions enhance the early detection of breast cancer and have wider implications for medical imaging, demonstrating the potential of AI in improving healthcare outcomes.

## List of Acronyms

|           |   |
|-----------|---|
| ACS       | American Cancer Society   |
| AGAN      | Autoencoder-Generative Adversarial Network                                  |
| AI        | Artificial Intelligence   |
| ANN       | Artificial Neural Network   |
| AUC       | Area Under Curve  |
| BC        | Breast Cancer   |
| BCDR      | Benchmark Breast Cancer Digital Repository                                  |
| BI-RADS   | Breast Imaging Reporting and Data System                                    |
| CAD       | Computer-Aided Diagnosis  |
| CAM       | Class Activation Maps   |
| CC        | Craniocaudal  |
| CBIS-DDSM | Curated Breast Imaging Subset of Digital Database for Screening Mammography |
| CLAHE     | Contrast Limited Adaptive Histogram Equalization                            |
| CNN       | Convolutional Neural Network  |
| COVID-19  | Corona Virus Disease of 2019  |
| COCO      | Common Objects in Context   |
| CT        | Computed Tomography   |
| DCGAN     | Deep Convolutional Generative Adversarial Network                           |
| DCNN      | Deep Convolutional Neural Network   |
| DDSM      | Digital Database for Screening Mammography                                  |
| DICOM     | Digital Imaging and Communications in Medicine                              |
| DCIS      | Ductal Carcinoma In Situ  |
| DL        | Deep Learning   |
| DNN       | Deep Neural Network   |
| HER       | Electronic Health Records   |
| FLOPS     | Floating Point Operations Per Second  |
| FID       | Fréchet Inception Distance  |
| FN        | False Negative  |
| FP        | False Positive  |
| GAN       | Generative Adversarial Network  |
| GAP       | Global Average Pooling  |
| HOG       | Histogram of Oriented Gradient  |

|         |   |
|---------|---|
| KNN     | K-Nearest Neighbors                               |
| LAG     | Large-Scale Attention-Based Glaucoma              |
| LBP     | Local Binary Pattern                              |
| m FPI   | mean False Positive Indications                   |
| MIAS    | Mammographic Image Analysis Society               |
| ML      | Machine Learning                                  |
| MLP     | Multilayer Perceptron                             |
| MLO     | Mediolateral Oblique                              |
| MO      | Mammographically-Occult                           |
| MRI     | Magnetic Resonance Imaging                        |
| MS-SSIM | Multiscale Structural Similarity Index Measure    |
| MVGG    | Modified Visual Geometry Group                    |
| NLP     | Natural Language Processing                       |
| NMS     | Non-Maximum Suppression                           |
| OMI-DB  | OPTIMAM Mammography Image Database                |
| PACS    | Picture Archiving and Communication System        |
| PET     | Positron Emission Tomography                      |
| PSNR    | Peak Signal-to-Noise Ratio                        |
| ReLU    | Rectified Linear Unit                             |
| RF      | Random Forest                                     |
| RCDT    | Radon Cumulative Distribution Transform           |
| RIM-ONE | Retinal Image Database for Optic Nerve Evaluation |
| RNN     | Recurrent Neural Network                          |
| SGD     | Sustainable Development Goals                     |
| ROI     | Region Of Interest                                |
| RSNA    | Radiological Society of North America             |
| SVM     | Support Vector Machine                            |
| TN      | True Negative                                     |
| TNM     | Tumor Node Metastasis                             |
| TP      | True Positive                                     |
| t-SNE   | t-distributed stochastic neighbor embedding       |
| UK      | United Kingdom                                    |
| UN      | United Nations                                    |
| USA     | United State of America                           |
| VGG     | Visual Geometry Group                             |

|         |                           |
|---------|---------------------------|
| ViT     | Vision Transformers       |
| WHO     | World Health Organization |
| XGBoost | Extreme Gradient Boosting |
| YOLO    | You Only Look Once        |

Table of Contents

Final Approval .....i

Declaration.....ii

Acknowledgments ..... iii

Dedication .....iv

ABSTRACT..... v

List of Acronyms .....vii

List of Tables .....xiv

List of Figures.....xv

Research Contribution .....xvii

Chapter 1 ..... 1

Introduction..... 1

1.1 Breast Cancer .....1

1.1.1 A Global Health Concern .....2

1.1.2 Importance of Early Prediction.....2

1.1.3 Breast Cancer Stages and BI-RADS Classification .....2

1.1.4 Conventional Methods for Identifying Breast Cancer.....5

1.2 Limitations of Conventional Techniques for Breast Cancer Detection .....7

1.3 Deep Learning in Medical Imaging .....8

1.3.1 Enhanced Diagnostic Capabilities .....9

1.3.2 Predictive Analytics and Risk Assessment.....9

1.3.3 Personalized Treatment Plans.....9

1.3.4 Drug Discovery and Development .....9

1.3.5 Potential Impact on Breast Cancer Diagnosis ..... 11

1.4 Challenges Faced by Deep Learning Models .....11

1.5 Research Gaps .....12

1.5.1 Diversity in Training Data ..... 12

1.5.2 Validation and Real-World Application..... 13

1.5.3 Minimizing Errors in Diagnosis ..... 13

1.6 Motivation .....13

1.7 Problem Statement.....14

1.8 Research Questions .....15

1.9 Aims and Objectives.....15

1.10 Scope of the Study .....16

1.11 Research Contribution.....17

1.12 Structure of the Thesis .....17

Chapter 2 .....19

Literature Review .....19

2.1 Data Augmentation .....19

2.1.1 Traditional Mammogram Augmentation Techniques .....21

2.1.2 GAN-based Mammogram Augmentation Techniques .....24

2.2 Synthetic Data Validation.....26

2.2.1 Visual Inspection .....27

2.2.2 Statistical Analysis .....28

2.3 Issues with Synthetic Data .....30

2.3.1 Lack of Realism.....30

2.3.2 Validation Challenges.....30

2.3.3 Bias and Overfitting.....31

2.4 Ensemble Classification of Images for Cancer Detection.....31

2.5 Visualizing Lesions in Breast Imaging .....38

Chapter 3 .....40

Synthetic Data Generation and Validation.....40

3.1 Introduction to Mammogram Generation and Validation .....40

3.1.1 The Imperative of Data Augmentation in Medical Imaging .....41

3.1.2 Data Scarcity in Breast Cancer Imaging.....42

3.1.3 Data Augmentation for Model Robustness and Generalization .....42

3.2 Methodology for Synthetic Data Generation and Validation .....44

3.2.1 Data Collection .....45

3.2.2 Data Preparation .....45

3.2.3 DCGAN Architecture .....46

3.2.4 Generator Network .....46

3.2.5 Discriminator Network .....47

3.2.6 Training Process .....48

3.2.7 Loss Function .....48

3.3 Performance Evaluation .....49

3.4 Experiments and Results .....50

3.3.1 Synthetic Mammogram Generation.....51

3.3.2 Mean Similarity .....61

3.3.3 Statistical Validation through Distance Calculation.....66

3.5 Validation of Synthetic Mammograms and Confidence Interval Analysis 69

3.6 Validation from Human Experts .....71

3.7 Chapter Summary.....71

Chapter 4 .....72

    Dual-View Mammogram Classification .....72

    4.1 Introduction to Dual Pathway Classification .....72

    4.2 Methodology for Dual Pathway EfficientViewNet Model .....74

        4.2.1 Dataset Description and Preprocessing .....75

        4.2.2 Architecture of the EfficientViewNet.....76

        4.2.3 CC View Pathway .....76

        4.2.4 MLO View Pathway.....79

    4.3 Evaluation Metrics .....81

        4.3.1.Confusion Matrix.....81

        4.3.2 Accuracy .....82

        4.3.3 Precision .....83

        4.3.4 Recall .....83

        4.3.5 F1 measure.....83

    4.4 Experiments and Results .....83

        4.4.1 Model Architecture and Training Setup .....83

        4.4.2 Results and Discussions.....84

        4.4.3 Comparative Analysis.....89

    4.5 Chapter Summary.....91

Chapter 5 .....92

Optimizing Breast Cancer Detection with Ensemble Learning .....92

    5.1 Introduction to Ensemble Model .....92

    5.2 Methodology for Ensemble Model.....94

        5.2.1 Data collection.....95

        5.2.2 Preprocessing and Augmentation of Data .....96

        5.2.3 Normalization and standardization of data .....96

        5.2.4 Model Selection and Architecture .....96

        5.2.5 EfficientNet Configuration .....97

        5.2.6 AlexNet configuration .....97

        5.2.7 ResNet Configuration .....98

        5.2.8 DenseNet Configuration .....98

        5.2.9 Model Training and Validation .....99

|  |            |
|--|------------|
| <b>5.3 Hyperparameter Optimization .....</b>             | <b>99</b>  |
| 5.3.1 Activation Function .....                          | 100        |
| 5.3.2 Training Process and Epoch Management .....        | 101        |
| 5.3.3 Validation Techniques and Overfit Prevention ..... | 101        |
| 5.3.4 Ensemble model development.....                    | 101        |
| 5.3.5 Model training .....                               | 103        |
| 5.3.6 Evaluation Metrics.....                            | 104        |
| <b>5.4 Results and Discussions .....</b>                 | <b>105</b> |
| 5.4.1. Performance of Individual Models .....            | 105        |
| 5.4.2 Ensemble Approach.....                             | 109        |
| 5.4.3 Comparative Analysis with Existing Models .....    | 112        |
| <b>5.5 Chapter Summary.....</b>                          | <b>112</b> |
| <b>Chapter 6 .....</b>                                   | <b>114</b> |
| <b>Lesion Visualization .....</b>                        | <b>114</b> |
| <b>6.1 Introduction .....</b>                            | <b>114</b> |
| 6.1.1 Importance of CAMs in Medical Imaging .....        | 114        |
| 6.1.2 Objectives of CAM in Lesions Analysis .....        | 115        |
| <b>6.2 Theoretical Background .....</b>                  | <b>116</b> |
| 6.2.1 CNN for Image Recognition.....                     | 117        |
| 6.2.2 Feature Activation in CNNs .....                   | 117        |
| 6.2.3 Evolution of CAM Techniques.....                   | 118        |
| <b>6.3 Results and Discussion .....</b>                  | <b>118</b> |
| 6.3.1 Quantitative Analysis .....                        | 120        |
| 6.3.2 Qualitative Analysis .....                         | 120        |
| 6.3.3 Assessment by Medical Experts .....                | 121        |
| <b>6.4. Discussion .....</b>                             | <b>122</b> |
| <b>6.5 Chapter Summary.....</b>                          | <b>123</b> |
| <b>Chapter 7 .....</b>                                   | <b>124</b> |
| <b>Conclusion and Future Work .....</b>                  | <b>124</b> |
| <b>7.1 Conclusion.....</b>                               | <b>124</b> |
| <b>7.2 Future Work .....</b>                             | <b>125</b> |
| <b>References .....</b>                                  | <b>127</b> |



List of Tables

Table 1.1: BI-RADS Classification ..... 4

Table 1.2: Limitations of Conventional Techniques for Breast Cancer Detection [16] ..... 8

Table 1.3: Advantages of the Deep Learning Techniques in Breast Cancer Detection [22] ... 10

Table 1.4: Challenges Faced by Deep Learning Model [26] ..... 11

Table 2.1: Comparison of Traditional Mammogram Augmentation Techniques ..... 23

Table 2.2: Comparison of GAN-based Mammogram Augmentation Techniques ..... 26

Table 2.3: Classification Techniques for Cancer Detection in Breast Images ..... 38

Table 3.1: Generator Network ..... 47

Table 3.2: Discriminator Architecture ..... 47

Table 3.3: Training Hyperparameters ..... 48

Table 4.1: CC View Pathway Configuration Parameters ..... 77

Table 4.2: MLO View Pathway Configuration Parameters ..... 79

Table 4.3: Performance Comparison with State-of-the-Art Techniques ..... 89

Table 5.1: Overall Performance of Models ..... 105

Table 5.2: Performance Comparison of Individual Models vs. Ensemble Approach ..... 110

Table 5.3: Comparative Analysis of Single Models vs. Ensemble Model ..... 111

Table 5.4: Performance Comparison with Existing Models ..... 112

## List of Figures

|  |    |
|--|----|
| Figure 1.1: An Overview of Breast Cancer Cell Sizes [11] .....                   | 5  |
| Figure 1.2 Stepwise Approach to Breast Cancer Detection.....                     | 5  |
| Figure 1.3: Mammography and Deep Learning.....                                   | 11 |
| Figure 2.1: Illustration of Traditional Mammogram Augmentation.....              | 22 |
| Figure 3.1: a) GAN-generated images, b) Real images .....                        | 43 |
| Figure 3.2: Methodology at Glance .....  | 45 |
| Figure 3.3: General Architecture of DCGAN.....                                   | 46 |
| Figure 3.4: Methodology for Validation of Synthetic Images .....                 | 49 |
| Figure 3.5: Synthesized images from epoch 0 during training of DCGAN.....        | 51 |
| Figure 3.6: Synthesized images from epoch 1 during training .....                | 52 |
| Figure 3.7: Synthesized images from epoch 2 .....                                | 53 |
| Figure 3.8: Synthesized images from epoch 3 during training .....                | 54 |
| Figure 3.9: Synthesized images from epoch 45 during training .....               | 55 |
| Figure 3.10: Epoch 50 Images .....   | 56 |
| Figure 3.11: Synthetic images during epoch 99.....                               | 57 |
| Figure 3.12: Images from epoch 100 during training .....                         | 58 |
| Figure 3.13: Discriminator and Generator Network Losses During Training .....    | 59 |
| Figure 3.14: Real vs Fake Losses During Training .....                           | 60 |
| Figure 3.15: Cluster Visualization of Synthetic Normal Class Total Data .....    | 61 |
| Figure 3.16: Cluster Visualization of Normal Class Data with Duplicate .....     | 62 |
| Figure 3.17: Cluster Visualization of Normal Class Validated Data .....          | 63 |
| Figure 3.18: Cluster Visualization of Benign Class Data with Duplicate .....     | 63 |
| Figure 3.19: Cluster Visualization of Synthetic Benign Class All Data.....       | 64 |
| Figure 3.20: Cluster Visualization of Synthetic Benign Validated Class Data..... | 64 |
| Figure 3.21: Cluster Visualization of Synthetic Malignant Class Data .....       | 65 |
| Figure 3.22 Visualization Malignant Class's Clusters with Duplicate .....        | 65 |
| Figure 3.23: Visualization of Malignant Validated Clusters .....                 | 66 |
| Figure 3.24: All Clusters Visualization of Synthetic Data with Duplicate .....   | 67 |
| Figure 3.25 All Clusters Visualization of Synthetic Data.....                    | 68 |
| Figure 4.1: Sample Mammograms from RSNA Benchmark Dataset.....                   | 74 |
| Figure 4.2 Methodology for Mammogram Classification .....                        | 75 |
| Figure 4.3: Architecture of the EfficientViewNet .....                           | 76 |
| Figure 4.4 Confusion Matrix.....   | 82 |
| Figure 4.5: Losses of the EfficientViewNet Model .....                           | 84 |

|   |     |
|---|-----|
| Figure 4.6: Training and Validation Accuracy of the EfficientViewNet.....             | 85  |
| Figure 4.7: Precision Status of the EfficientViewNet Model .....                      | 86  |
| Figure 4.8: Recall of the EfficientViewNet Model.....                                 | 87  |
| Figure 4.9: F1 Score of the Model .....   | 87  |
| Figure 4.10: Optimization of Learning Rate in EfficientViewNet .....                  | 88  |
| Figure 4.11: EfficientViewNet Vs. State-of-the-Art.....                               | 90  |
| Figure 4. 12: Performance Comparison of EfficientViewNet vs DenseNet & ResNet Set ... | 91  |
| Figure 5.1: Methodology of Ensemble Model.....  | 95  |
| Figure 5.2: Basic Architecture of EfficientNet .....                                  | 97  |
| Figure 5.3: Architecture of AlexNet .....   | 97  |
| Figure 5.4: Architecture of AlexNet .....   | 98  |
| Figure 5.5: DenseNet Architecture .....   | 99  |
| Figure 5.6: Ensemble Layout.....  | 102 |
| Figure 5. 7 Confusion Metrix for EfficientNet Model.....                              | 106 |
| Figure 5. 8: ResNet Model Confusion Matrix .....                                      | 107 |
| Figure 5. 9: Confusion Matrix for AlexNet model .....                                 | 108 |
| Figure 5. 10: Confusion Matrix for DenseNet model .....                               | 108 |
| Figure 5.11: Accuracy of Models .....   | 109 |
| Figure 5. 12: Performance Metrics of ensemble vs. individual models.....              | 110 |
| Figure 5.13: Comparative Analysis of Single Models vs. Ensemble Model .....           | 111 |
| Figure 5.14: Sensitivity and Specificity Comparison .....                             | 111 |

## Research Contribution

The following Research Papers related to this thesis are Published/submitted to journals during the PhD research.

### International Journal Publication:

Chapter 3 details the methodology of using DCGAN for generating high-quality synthetic mammograms, addressing the challenge of limited data. The techniques for validating these synthetic images to ensure data integrity and quality are also thoroughly discussed.

#### Published Papers

1. Dilawar Shah, Mohammad Asmat Ullah Khan, Mohammad Abrar. "Reliable Breast Cancer Diagnosis with Deep Learning: DCGAN-Driven Mammogram Synthesis and Validity Assessment." *Applied Computational Intelligence and Soft Computing*, 2024. **(Published IF 2.4)**
2. Dilawar Shah, Mohammad Asmat Ullah Khan, Mohammad Abrar, et al. "Enhancing the Quality and Authenticity of Synthetic Mammogram Images for Improved Breast Cancer Detection." *IEEE Access*, 2024. **(Published IF 3.4)**

EfficientViewNet architecture effectively handles modality shifts by leveraging two pathways in Chapter 4. It enhances model performance through detailed feature extraction from CC and MLO mammographic views is extensively covered.

#### Submitted Papers

1. Dilawar Shah, Mohammad Asmat Ullah Khan, Mohammad Abrar, Muhammad Tahir. "Dual-View Deep Learning Model for Accurate Breast Cancer Detection in Mammograms" *International Journal of Intelligent Systems*, 2024.  
Under Review: First Round Review Submitted **[IF: 5.0]**

Chapter 5 focuses on the development of an ensemble model combining EfficientNet, DenseNet, ResNet, and AlexNet to improve classification accuracy.

#### Submitted Papers

1. Dilawar Shah, Mohammad Asmat Ullah Khan, Mohammad Abrar, Muhammad Tahir. "Optimizing Breast Cancer Detection with an Ensemble Deep Learning Approach" *International Journal of Intelligent Systems*, 2024.  
Accepted **[IF: 5.0]**

# Chapter 1

## Introduction

In the contemporary digital era, the accelerated progression of information technology has substantially transformed multiple sectors, with healthcare being a primary beneficiary of this revolution. Innovations such as electronic health records (EHRs) and artificial intelligence (AI)-powered decision support systems have not only optimized healthcare delivery processes but have also significantly increased the accuracy of clinical decision-making. By 2020, projections suggested that the digitization of healthcare would result in the creation of approximately 2314 exabytes of data. Although this immense volume of data presents promising opportunities for advancements in medical research and patient care, it also introduces considerable challenges in the efficient analysis and extraction of actionable insights, which are critical for improving diagnostic precision and treatment outcomes[1-5].

### 1.1 Breast Cancer

Breast cancer (BC), a formidable adversary, silently claims the lives of thousands of women each year. It is a severe disease among women, especially in the age group of 40 to 55 years. Breast cancer is one of the deadliest types of cancer after lung cancer. Globally, over 2.3 million cases were diagnosed, and more than 508,000 lives were lost only in 2020. American Cancer Society (ACS) assembles the latest records on cancer prevalence, mortality, and survival yearly [6]. A tumor or lumps could be malignant or benign. Benign is a rare feast to the other parts of the body and is often considered non-cancerous. If these lumps are not creating a problem for the women, like breast pain or spreading to the other tissues, then they are ignored. Benign lumps may be cysts, phyllode tumors, atypical hyperplasia, fibroadenomas, adenosis, and fat necrosis. Malignant tumors are invasive and can spread to other parts of the body, like the liver, bones, brain, or lungs. If these invasive tumors are not detected well in time, they destroy the encompassing breast tissue and can lead to metastatic breast cancer[7].

### **1.1.1 A Global Health Concern**

Breast cancer is a significant worldwide health issue that affects people, families, and societies across different regions. This condition affects women but can also impact men, making it a matter of universal significance. It is a highly prevalent cancer worldwide, with millions of new cases reported each year. This high prevalence underscores the importance of understanding and addressing disease globally. Beyond its prevalence, breast cancer's impact is felt through its associated morbidity and mortality rates. It is not only the most common cancer among women but also a leading cause of cancer-related deaths globally. Breast cancer not only affects physical health but also has profound psychosocial and economic consequences. It can create a substantial emotional toll on individuals and families, leading to anxiety and emotional distress. Additionally, the cost of diagnosis, treatment, and potential work-related disruptions can have wide-ranging economic effects [8].

### **1.1.2 Importance of Early Prediction**

The gravity of breast cancer's global impact underscores the significance of early detection. Timely diagnosis can lead to more effective treatment and better survival rates. Given its extensive impact, breast cancer research and awareness campaigns have become critical components of the global health agenda. Researchers, healthcare professionals, and advocacy groups worldwide are working diligently to understand the disease better, develop effective treatments, and promote early detection [9].

### **1.1.3 Breast Cancer Stages and BI-RADS Classification**

In breast cancer diagnosis, two major systems are used to assess the severity of the disease: the TNM (Tumor, Node, Metastasis) staging system and the BI-RADS classification system. Each plays a critical role in helping clinicians understand the progression of cancer and determine appropriate treatment strategies. In the TNM system categorizes cancer based on three key factors including Tumor (T) which shows the size and extent of the primary tumor. Node (N) which shows whether the cancer has spread to nearby lymph nodes. Similarly, Metastasis (M) signifies whether the cancer has spread to other parts of the body.

Based on these criteria, breast cancer is grouped into Stages 0 through IV. Stage 0 refers to Ductal Carcinoma In Situ (DCIS), a non-invasive form where abnormal cells are

confined to the milk ducts. Stages I and II are early stages where the tumor remains localized to the breast or nearby lymph nodes and is typically smaller in size (up to 5 cm). Stage III indicates a larger tumor or one that has spread more extensively into the lymph nodes, skin, or chest wall. Stage IV is the most advanced stage, where cancer has metastasized to distant organs such as the liver, lungs, or bones [10]. The Breast Imaging Reporting and Data System (BI-RADS) is used to standardize the interpretation of imaging results from mammograms, ultrasounds, and Magnetic Resonance Imaging MRIs. It helps radiologists describe what they see and guide the next steps in a patient's care. BI-RADS uses a numerical system from 0 to 6, which reflects the likelihood that an abnormality is malignant. The BI-RADS system serves as a critical tool, helping radiologists communicate findings clearly to oncologists and ensure timely follow-ups. It bridges the gap between imaging results and clinical action, whether the findings indicate benign changes or require further intervention.

As breast cancer progresses from early to later stages, the appearance on mammograms changes. In early-stage cancer (Stages 0-II), mammograms may reveal small, localized masses or microcalcifications, which are tiny deposits of calcium that appear as white spots on the image. These signs can be subtle, requiring careful analysis to detect, especially when tumors are small (less than 2 cm). In later stages (Stages III-IV), the tumors are larger, and mammograms often show more pronounced abnormalities, including irregular masses and architectural distortion of breast tissue. At these stages, the likelihood of cancer spreading to lymph nodes or distant organs is higher, and the abnormalities on the images are more noticeable.

There is a strong correlation between the stages of cancer (as per TNM) and the BI-RADS categories. BI-RADS categories 3 and 4 often correspond to early-stage cancer (Stages I and II), where findings are suspicious but not conclusive. BI-RADS categories 5 and 6, on the other hand, are more likely to be associated with later-stage cancer (Stages III and IV), where the likelihood of malignancy is significantly higher.

By using both the TNM staging and BI-RADS classification systems, clinicians can gain a comprehensive understanding of a patient's breast cancer status, from the size and spread of the tumor to the likelihood of malignancy based on imaging results. This combined approach ensures that patients receive timely and accurate diagnoses, guiding

treatment plans that are best suited to their specific condition. Level 6 indicates a high cancer level, as shown in Table 1.1.

**Table 1.1: BI-RADS Classification**

| Category |   | Description                    | Likelihood of cancer | Recommendations                 |
|----------|---|--------------------------------|----------------------|---------------------------------|
| 0        |   | Inconclusive                   | n/a                  | Recall for additional Screening |
| 1        |   | Normal                         | 0%                   | Routine screening               |
| 2        |   | Benign finding                 | 0%                   | Routine screening               |
| 3        |   | Benign finding                 | > 0% but $\leq$ 2%   | Short term follow-up (6 months) |
| 4        |   | Abnormality requiring a biopsy | > 2% to < 95%        | Biopsy                          |
|          | A | Low suspicion                  | > 2% to $\leq$ 10%   | Biopsy                          |
|          | B | Moderate suspicion             | > 10% to $\leq$ 50%  | Biopsy                          |
|          | C | High suspicion                 | > 50% to < 95%       | Biopsy                          |
| 5        |   | Highly suspected cancer        | $\geq$ 95%           | Biopsy                          |
| 6        |   | Biopsy-proven malignancy       | 100%                 | Surgical procedure              |

Breast cancer can be divided into early (stages 0-II) and later stages (stages III-IV), each presenting distinct mammographic characteristics. In stages 0 to II, mammograms may show tiny, localized masses or microcalcifications. Calcification is calcium deposits that appear as white specks. Tumors are often smaller than 2 cm, making them harder to detect but more responsive to treatment. In advanced stages (III and IV), mammographic images often reveal larger, irregular masses that may have invaded nearby tissues. Tumors are more noticeable as they cause architectural distortion of the breast tissue. In Stage IV, cancer has metastasized distant organs, with significant changes visible in the breast. BI-RADS categories 3 and 4 typically correspond to early-stage cancers, where findings are suspicious but not conclusive. These categories often require short-term follow-ups or biopsies. BI-RADS categories 5 and 6, on the other hand, are associated with later-stage cancer, where the likelihood of malignancy is significantly higher and requires immediate intervention.

This standardized reporting system enables timely and precise diagnosis, allowing patients to receive appropriate treatments based on the stage and severity of their breast cancer. Mehrotra *et al.* [11] reported that a tumor is detectable on a mammogram after almost 8 years of its inception. Cancer cells are doubled every 90 days, so there are



more than 4 billion cancer cells when the tumor is detectable. A comparison is presented about the size of the cancer cell is depicted in Figure 1.1.

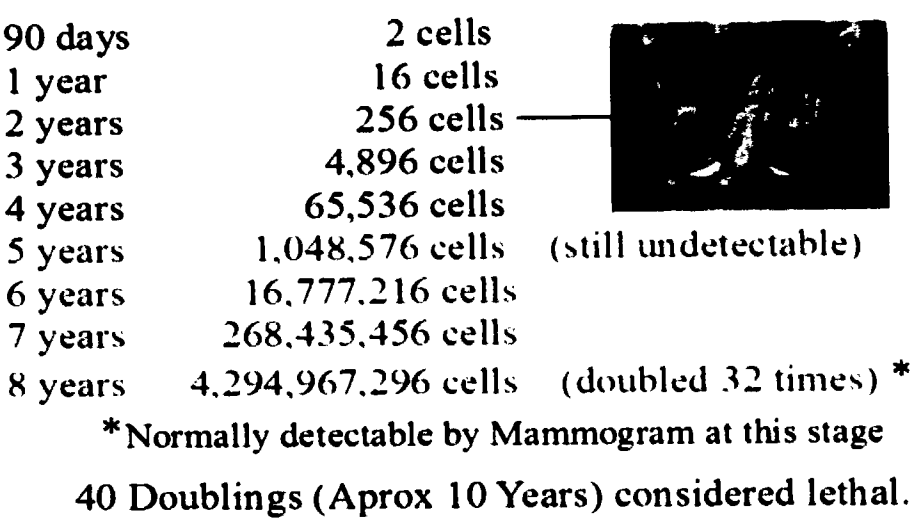


Figure 1.1: An overview of Breast Cancer Cell Sizes [11]

1.1.4 Conventional Methods for Identifying Breast Cancer

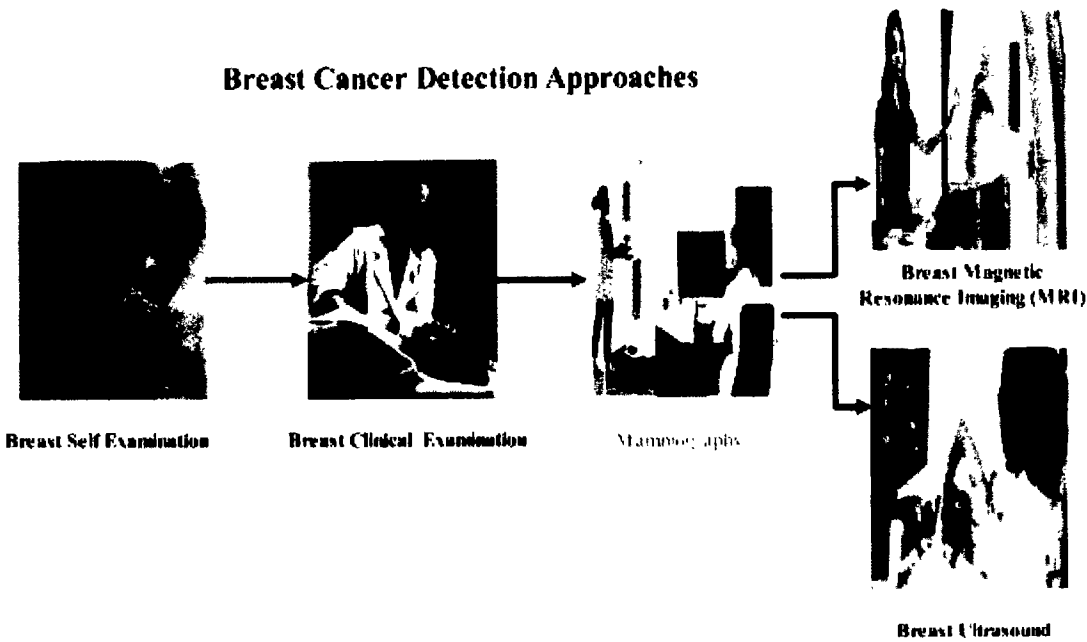


Figure 1. 2 Stepwise Approach to Breast Cancer Detection

This diagram in Figure 1. 2 illustrates the standard diagnostic workflow for breast cancer detection. It begins with Breast Self-Examination, where individuals monitor for early signs or symptoms. If abnormalities are detected, the next step is a Clinical Breast

Examination conducted by a healthcare professional. Upon clinical suspicion, Mammography is performed to obtain detailed imaging of the breast tissue. If the mammogram indicates potential malignancy, advanced diagnostic modalities such as Breast MRI or Breast Ultrasound are employed for further evaluation and confirmation.

Radiologists in daily practice are using different imaging techniques for breast cancer detection. The most commonly used techniques are mammograms, followed by ultrasound and MRI. However, mammography is commonly considered the most common and well-established method for early cancer screening of the breast. Mammography involves using low-dose X-rays to generate mammographic images of the breast tissue. The images effectively detect tumors that may not be palpable during physical examinations. However, there are limitations, including the possibility of false positives and false negatives [12]. Ultrasound of the breasts uses high-frequency sound waves to generate images of the tissues in the breast. Ultrasound is often used as a follow-up to mammography to evaluate abnormalities further. Ultrasound is beneficial in differentiating between solid masses, which may be cancerous, and fluid-filled cysts, usually benign tissues. However, it is less effective in detecting microcalcifications, which can be an early sign of breast cancer [13]. On the other hand, Breast MRI is a powerful imaging technique that can provide detailed images of breast tissue. It is often used in cases where mammography or ultrasound results are inconclusive or for screening high-risk individuals. Breast MRI is highly sensitive but is not typically used as a first-line screening method due to its cost and the need for intravenous contrast. [14]. Breast Thermography, also known as thermal imaging of the breast, is a non-invasive imaging technique that uses infrared technology to detect and visualize temperature variations in the breast tissue. It is non-invasive and does not involve any contact with the breast tissue. However, it is not considered a standalone method for diagnosing breast cancer. Some healthcare professionals and organizations are cautious about its reliability and recommend traditional breast screening methods as the primary means of early detection.

### **1.1.5 Advantages of Conventional Techniques for Breast Cancer Detection**

Traditional diagnosing techniques for breast cancer, such as mammography, ultrasound, and clinical breast examination, have been thoroughly tested and proven effective over many years. These techniques offer dependable early-stage cancer

detection, key to reducing mortality and enhancing patient outcomes. They are readily available in a wide range of healthcare settings, making them accessible to a large population segment. One of their main strengths lies in the expertise of healthcare professionals who have invaluable experience and skill in interpreting results. This human element adds a layer of nuanced evaluation and patient-specific care vital in the diagnostic process. Moreover, these traditional methods follow established protocols and guidelines, ensuring consistency, quality control, and standardized practices across different healthcare facilities. This standardization also aids in minimizing variability in diagnoses from one observer to another and supports continuous improvement in diagnostic techniques. Another significant aspect of these techniques is their ability to monitor breast health over time. Regular screenings enable the detection of minor changes, thus increasing the chances of early diagnosis and reducing missed cases. The results from these methods are easy to interpret, allowing healthcare providers to make well-informed decisions and discuss treatment options clearly with patients [15]. This clarity in results is crucial for crafting individualized treatment plans based on each patient's unique case. Regarding cost, traditional diagnostic methods like mammography are economical, especially compared to some of the more advanced imaging or genetic testing options. This cost-effectiveness makes them a viable option for large-scale screening programs and in settings where resources are limited.

## **1.2 Limitations of Conventional Techniques for Breast Cancer Detection**

Over time and with advancements in technology, breast cancer screening has also progressed. Nevertheless, the analysis of mammographic images poses a formidable challenge. Various factors, such as the quality of the image and the level of expertise of the radiologist influence visual perception. The cost and quality of mammograms vary. The level of expertise of senior radiologists surpasses that of juniors. Empirical evidence shows that in real-world scenarios, mammographers may fail to detect 10-30 percent cancers, resulting in what is known as a False Negative. Conversely, 65 to 85 percent of women who undergo mammography are summoned for a repeat procedure, known as false positives. In such circumstances, individuals may even be required to undergo an unwarranted biopsy [16]. In either case, the occurrence of high false positive rates and false negatives leads to overdiagnosis, overtreatment, and unnecessary biopsies for the patients. The main limitations are summarized in Table 1.2.

**Table 1.2: Limitations of Conventional Techniques for Breast Cancer Detection [16]**

| S. No | Limitation                           | Explanation  |
|-------|--------------------------------------|--|
| 1     | Higher Subjectivity and Variability  | Dependence on healthcare professionals' judgment leads to diagnosis variability influenced by their differing skill levels and experience.                                 |
| 2     | Incorrect Diagnoses and Missed Cases | Potential for incorrect results, where benign conditions are identified as cancer (false positives) or cancerous conditions are missed (false negatives).                  |
| 3     | Low Sensitivity for Dense Breasts    | Mammography's diminished effectiveness in dense breast tissue can hide smaller tumors, risking missed diagnoses and treatment delays.                                      |
| 4     | High Risk                            | Biopsies, necessary for a definitive diagnosis, are invasive, can cause discomfort, have complication risks, and involve additional costs.                                 |
| 5     | Restricted Reach                     | Higher costs and lower availability of some techniques, like breast MRI, limit access, especially in resource-limited settings.  |
| 6     | Limited Forecasting Insight          | Molecular diagnostics and genetic testing may not provide a complete picture of cancer's behavior and prognosis, necessitating further information for treatment planning. |

To address such potential diagnosis errors, a double reading is an efficient solution to minimize the number of false positive or false negative cases. In real life, double reading is not feasible due to the lack of expert radiologists and additional patient costs. Computer Aided Diagnosis (CAD) is reasonable assistance for radiologists in making decisions. This adoption is a good and affordable solution. As a second reader, CAD encourages the radiologist to reexamine suspicious cases again.

In reality, CAD has shown promising performance as a double reader, but the final decision remains with the radiologists.

**1.3 Deep Learning in Medical Imaging**

DL has emerged as one of the most transformative forces in healthcare, offering new avenues to enhance patient care, streamline healthcare operations, and drive medical research. Its role in healthcare is multifaceted and holds significant promise for the future of medicine.

### 1.3.1 Enhanced Diagnostic Capabilities

DL is progressively employed to analyze medical images, including interpreting X-rays, MRIs, and Computed Tomography (CT) scans. These algorithms can accurately detect patterns and irregularities in medical images, aiding radiologists in timely identifying diseases such as cancer and diminishing the chances of incorrect diagnosis [17]. Moreover, Natural language processing (NLP) methods can analyze unstructured clinical notes, helping doctors extract information from EHRs. This assists in diagnosing and managing patients more effectively [18].

### 1.3.2 Predictive Analytics and Risk Assessment

DL models can analyze patient data to identify high-risk individuals for specific diseases or adverse events. These models help healthcare providers offer initiative-taking care and interventions, improving patient outcomes. Machine learning (ML) algorithms can analyze vast datasets to predict disease outbreaks and public health trends, which is particularly valuable in managing epidemics and resource allocation [19].

### 1.3.3 Personalized Treatment Plans

DL can analyze a patient's genetic information and medical history to tailor treatment plans. This level of personalization ensures that treatments are practical and minimize side effects [20].

### 1.3.4 Drug Discovery and Development

DL is revolutionizing drug discovery by significantly speeding up the process. ML models can analyze vast chemical and biological datasets to identify potential drug candidates, predict their efficacy, and optimize molecular structures. DL and GANs generate molecular structures and predict how they interact with biological targets, saving time and resources in developing new drugs [21].

It can optimize hospital operations, from scheduling appointments and managing patient flow to resource allocation. This not only reduces administrative burden but also enhances patient experience. Robotics and automation are used in surgical procedures, reducing the margin of error and allowing for minimally invasive surgeries [22]. The Advantages of the DL techniques are summarized as follows in Table 1.3.

**Table 1.3: Advantages of the Deep Learning Techniques in Breast Cancer Detection [22]**

| S. No | Advantage                              | Explanation  |
|-------|--|--|
| 1     | Comprehensive Data Analysis            | These algorithms are proficient in processing and analyzing large and complex datasets, including clinical records, imaging, genetic data, and patient demographics.   |
| 2     | Identifying Patterns and Anomalies     | Such methods excel at recognizing intricate patterns and identifying subtle anomalies that doctors might overlook.   |
| 3     | Increased Diagnostic Accuracy          | It can enhance diagnostic precision by reducing human error and minimizing differences in interpretations among clinicians.  |
| 4     | Integrating Varied Data Sources        | It efficiently combine patient data from multiple sources, such as clinical records, imaging, and genetics.  |
| 5     | Tailored Risk Assessment and Treatment | It can create individualized risk assessment models, aiding healthcare professionals in customizing treatment plans.   |
| 6     | Flexibility and Learning Capacity      | These approaches are adaptable to varying data sets and can learn from extensive data volumes.   |
| 7     | Facilitating Early Detection           | These models have the potential to play a crucial role in the early detection of breast cancer. Identifying these early indicators allows for timely intervention, which can enhance treatment success rates and improve patient survival. |
| 8     | Comprehensive Data Analysis            | They are proficient in processing and analyzing large and complex datasets, including clinical records, imaging, genetic data, and patient demographics.   |

To address such potential diagnosis errors, a double reading is an efficient solution to minimize the number of false positive or false negative cases. In real life, double reading is not feasible due to the lack of expert radiologists and additional patient costs. CAD is a reasonable assistance for radiologists in making decisions. This adoption is a good and affordable solution. As a second reader, CAD encourages the radiologist to reexamine suspicious cases again. In reality, CAD has shown promising performance as a double reader, but the final decision remains with the radiologists. The advent of these techniques provides a ray of hope in the detection of breast cancer. DL has demonstrated remarkable capabilities in recognizing complex patterns within medical images, making it particularly well-suited for analyzing mammograms and histopathological images. In the broader context, DL models, such as CNNs, can discern subtle abnormalities in medical images, potentially reducing the occurrence of

false positives and false negatives in medical settings. Such innovative models can be fine-tuned to enhance their performance in breast cancer detection in real life [23].

1.3.5 Potential Impact on Breast Cancer Diagnosis

The potential impact of this research extends beyond theoretical advancements. DL has a promising role in mammography, as elaborated in Figure 1.3.

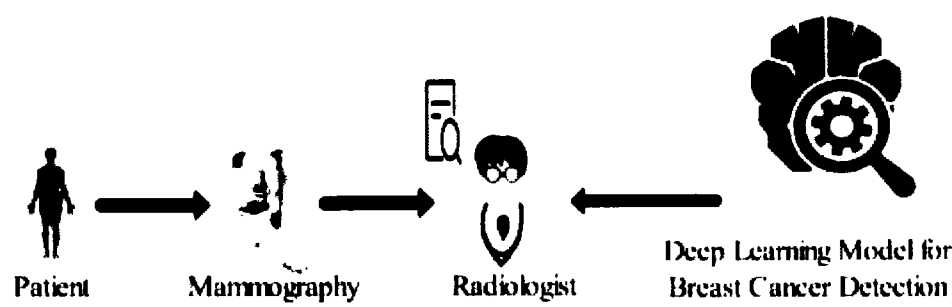


Figure 1.3: Mammography and Deep Learning

It is anticipated that the integration of DL, GANs, and Class Activation Maps (CAMs) will significantly enhance the accuracy and efficacy of breast cancer diagnosis and classification [24]. The success of AlexNet [25], and other recent architectures on the ImageNet challenge popularized it. Researchers use these base architectures for mammographic classification and other breast cancer-related tasks.

1.4 Challenges Faced by Deep Learning Models

DL algorithms hold significant promise for advancing breast cancer diagnosis, but there are key challenges that must be considered and are detailed in Table 1.4

Table 1.4: Challenges Faced by Deep Learning Model [26]

| S. No | Challenge                  | Description   |
|-------|----------------------------|---|
| 1     | Dependence on Data Quality | The efficacy of these models is intricately linked to the caliber and variety of the data they are trained on. Insufficient or prejudiced data can result in erroneous forecasts. Inadequate diversity in the training dataset can lead to suboptimal generalization, thereby impacting accuracy across various populations or subgroups. |
| 2     | Interpretability           | DL models are often seen as "black boxes." They might deliver accurate results but lack clear explanations for their decisions, hindering their acceptance and trustworthiness in clinical  |

| S. No | Challenge                          | Description  |
|-------|------------------------------------|--|
|       |                                    | settings. Ongoing research is attempting to make these methods more interpretable.   |
| 3     | Risk of Overfitting                | Overfitting can occur when the model is excessively tailored to the training data, resulting in reduced performance when applied to new, unseen data. Overfitting can lead to high accuracy during the training phase, but it often fails to generalize well in real-world situations. To tackle this issue, it is necessary to implement suitable validation, regularization, and hyperparameter tuning techniques. |
| 4     | Challenges in Data Acquisition     | Developing these algorithms often requires extensive, high-quality labeled data, which can be difficult to obtain, particularly for rare cancer types or specific subtypes. This limitation can impact the 5-algorithm's performance and applicability in specialized diagnostic contexts.   |
| 5     | Ethical and Privacy Concerns       | Applying DL in medical diagnosis involves crucial ethical issues, including ensuring patient privacy, data confidentiality, and security. Additionally, there is a need to address potential biases in the algorithms to prevent exacerbating existing health disparities.   |
| 6     | Integration into Clinical Practice | To effectively use such algorithms in healthcare in general, particularly breast imaging, they must be integrated seamlessly into existing clinical workflows. This involves overcoming technical challenges, ensuring compatibility with current systems, and providing interfaces healthcare professionals can easily use to interpret the algorithm's results.  |

1.5 Research Gaps

In the realm of DL for breast cancer diagnosis, significant progress has been achieved. Yet, to fully realize the abilities of these promising technologies, various research gaps must be addressed. Addressing these research gaps is vital for advancing the application of DL in breast cancer diagnosis and treatment, ensuring its efficacy, reliability, and ethical alignment with patient care standards. Some gaps are presented below.

1.5.1 Diversity in Training Data

The diversity of training data is fundamental to DL algorithms, which directly influences pattern recognition and predictive accuracy. Current datasets often lack representation across different populations, ethnic groups, and breast cancer subtypes, leading to potential biases and limited algorithmic applicability [25].



### 1.5.2 Validation and Real-World Application

Despite promising research outcomes, the real-world clinical validation of DL algorithms in breast cancer diagnosis remains nascent. There is a pressing need for extensive clinical trials to assess these algorithms' practical utility and seamless integration into existing healthcare workflows [38].

### 1.5.3 Minimizing Errors in Diagnosis

False positives and false negatives have critical ramifications for patient care in the context of breast cancer. A false positive, where noncancerous conditions are mistakenly identified as cancer, can lead to unnecessary stress, additional testing, and sometimes needless treatments, which can be both physically and emotionally taxing for patients. On the other hand, false negatives, which occur when actual cases of cancer are missing, can result in delayed treatment with potentially detrimental effects on patient outcomes. It is vital to minimize both to improve patient care quality and reduce healthcare costs [27].

## 1.6 Motivation

The impetus for this research is deeply rooted in the pressing challenges and transformative opportunities within the landscape of medical diagnostics, particularly in breast cancer detection. Central to this pursuit is the critical issue of data scarcity in medical research, exacerbated by the sensitive nature of patient information. In breast cancer diagnosis, the limited availability of annotated training data presents a formidable obstacle, impacting the development and efficacy of DL models. This research endeavor, therefore, emerges from the need to transcend these constraints, leveraging innovative techniques to enhance the dataset's richness and reliability.

The study begins by confronting the enduring challenge of developing DL models capable of learning from minimal data. Traditional data augmentation methods, such as flipping, rotating, and mirroring images, often steer models towards overfitting, as they tend to memorize features from repetitively altered images. This research proposes to navigate beyond this limitation, utilizing DCGAN to generate high-quality synthetic

mammograms. This approach balances the dataset, enriching it with diverse and representative samples to train more robust and accurate models.

However, the generation of synthetic data using GANs has its pitfalls. The nuanced balance between making these images realistically diverse and avoiding homogeneity is crucial to avoid diagnostic inaccuracies. Recognizing this, the research introduces a rigorous validation mechanism to ensure that each synthetically generated mammogram accurately reflects the class it represents, thereby preserving the integrity and utility of the augmented dataset.

Another focal point of this study is the optimization of CNN architectures to enhance mammogram classification accuracy. Addressing the common pitfalls of misclassification, such as false positives and negatives, is paramount in a domain where accuracy is not just a metric but a determinant of patient outcomes. By developing an ensemble model that harnesses the strengths of various CNN architectures, the research aims to significantly reduce misclassification rates, contributing to a more reliable and early diagnosis of breast cancer.

Finally, recognizing the gap between advanced computational models and their interpretability by medical professionals, this study endeavors to bridge this divide. By employing class activation maps, the research seeks to render the decision-making process of DL models transparent, visually highlighting regions of interest such as masses or defective cells. This approach not only aids in validating the model's predictions but also makes its insights accessible and understandable to domain experts.

## **1.7 Problem Statement**

In medical research, particularly breast cancer diagnosis, the scarcity of annotated data is a critical challenge due to the sensitivity of patient information. Although numerous deep-learning (DL) models have been developed to predict breast cancer using limited datasets, these models struggle to achieve robustness, which typically requires extensive data for training. Current data augmentation techniques, such as flipping, rotating, and mirroring, while helpful, often lead to overfitting as models memorize features from repeated images. Additionally, generating synthetic data through methods like GANs poses the risk of producing images that are either too similar or not similar

enough to real medical images, which could lead to diagnostic inaccuracies without proper validation. Furthermore, despite the potential of CNNs in medical image analysis, they are prone to misclassification errors, manifesting as false positives and false negatives, which can severely impact patient outcomes. Lastly, CNN models lack interpretability, preventing domain experts from understanding how predictions are made and failing to highlight critical regions, such as masses or defective cells, in medical images, limiting their practical utility in clinical settings.

## 1.8 Research Questions

The following questions arise from the detailed discussion of the problem statements.

1. How effectively can synthetic data augmentation address mammograms' data scarcity issue?
2. How is an effective validation mechanism devised to check whether the augmented mammograms are accurate according to the class from which they are created?
3. How can various CNN architecture (s) be utilized to improve the accuracy of the mammogram classification?
4. How do we explain the results of CNN through visualization to make it more understandable for domain experts?

## 1.9 Aims and Objectives

This research aims to enhance the accuracy and interpretability of breast cancer detection using DL by addressing the challenges of data scarcity and misclassification.

The objectives of this research study are elaborated as follows:

1. To balance the mammogram dataset using augmentation techniques.
2. To validate and annotate the augmented dataset.
3. To develop a robust ensemble model to improve breast cancer detection classification accuracy using CNN.
4. To visualize and explain the defective cells using a class activation map.

## 1.10 Scope of the Study

This research aligns with the United Nations' (UN) Sustainable Development Goals, particularly SDG 3: Good Health and Well-being. By addressing the critical challenge of early breast cancer detection, our work contributes to the broader objective of reducing the global health burden of cancer. Leveraging advanced computational techniques to improve diagnostic accuracy directly aids in elevating the quality of healthcare, potentially reducing mortality rates associated with breast cancer and contributing to the attainment of the SDG target of ensuring healthy lives and promoting well-being for all at all ages.

This research introduction highlights the significant capacity for transformation in medical imaging and oncology. Using DCGAN to expand the dataset tackles the common problem of insufficient data in medical research, which frequently impedes the creation of strong deep-learning models. Our methodology creates high-fidelity mammographic images, increasing the data available for training and improving the diversity and representativeness of existing datasets. Moreover, the verification and improvement of the produced images using an innovative approach guarantee the dependability of the data, promoting confidence in AI systems among healthcare practitioners. The subsequent classification process aims to minimize false positives and negatives can significantly improve patient outcomes, reduce unnecessary biopsies, and streamline the diagnostic pathway, thereby enhancing healthcare services' overall efficiency and usefulness.

The implications of our research for society are manifold. Firstly, it empowers radiologists and oncologists with state-of-the-art tools for breast cancer detection, thereby enhancing the diagnostic process and enabling more timely and accurate interventions. This, in turn, can lead to improved survival rates, reduced treatment costs, and less emotional strain on patients and their families. Using CAMs for visualization further demystifies the decision-making process of neural networks, making the technology more transparent and interpretable for domain experts by providing visual explanations for model predictions.

Our study is positioned at the intersection of innovation and societal well-being. It exemplifies the profound impact that thoughtful integration of AI can have on public health initiatives, healthcare delivery, and the broader societal quest for sustainable

development. As we continue to refine and expand upon this work, it promises to set new benchmarks in healthcare technology and contribute to a healthier society.

### 1.11 Research Contribution

1. Balanced the mammogram dataset through DCGAN to generate high-quality and diverse class-wise synthetic mammograms, addressing the challenge of limited data.
2. Validated the augmented dataset using similarity assessment and standard deviation methods, ensuring data integrity and quality. We devised a mechanism to identify outliers and redundant images in each class. To refine the dataset, these images were removed from the synthetic dataset. We also validated the images through radiologists.
3. Developed EfficientViewNet architecture, effectively mitigating modality shift by leveraging two pathways. The distinctive architecture consistently enhances the performance of the model. Moreover, developed a robust ensemble model combining EfficientNet, DenseNet, ResNet, and AlexNet to improve classification accuracy in breast cancer detection using CNNs.
4. Employed CAMs to visualize and interpret cancerous cells, enhancing the comprehensibility of results for medical experts

### 1.12 Structure of the Thesis

The structure of the thesis is organized as follows:

Chapter 1 introduction to breast cancer, breast imaging modalities, breast cancer stages, breast cancer stages, and BI-RADS classification, the role of DL in healthcare and imaging, challenges in DL models, objectives, problem statement, and research questions.

Chapter 2 presents and summarizes the literature review and findings of the various state-of-the-art techniques for mammographic data augmentation, validation, classification, and visualization. These techniques are assessed in various parameters, and other factors are further used for breast cancer detection.

Chapter 3 describes the details of the DCGAN-based data augmentation methodology. Data scarcity and need for data augmentation in breast cancer imaging, DCGAN architecture, training process of DCGAN network, performance evaluation of DCGAN generated images, synthetic mammogram generation, clusters visualization of classes.

Chapter 4 discusses the classification of mammograms for breast cancer detection. The EfficientViewNet architecture has two parallel pathways. The MLO and CC view pathways are intended to accommodate MLO and CC view mammograms, respectively. These pathways effectively extracted more relevant and detailed features from these mammographic views, improving model performance. The results of the EfficientViewNet model are compared to the state of the art at the end of the chapter.

Chapter 5 introduces an ensemble methodology for classifying breast cancer. The ensemble models integrate innovative models, including AlexNet, DenseNet, EfficientNet, and ResNet. The chapter thoroughly examines and compares these classifiers, emphasizing their individual and group effectiveness in detecting breast cancer. The ensemble approach combines the predictive abilities of individual models, resulting in higher accuracy and highlighting the beneficial collaboration of combined classifier knowledge.

Chapter 6 presents the importance of CAMs in medical imaging, the objectives of CAM in breast cancer detection, the background of CAM in medical imaging, the evolution of CAM techniques, CAM visualization, and its quantitative and qualitative analysis and validation from human experts.

Chapter 7 presents the conclusion and future work. Future research will focus on advancing GAN architectures and integrating them with other ML techniques for better mammogram generation, alongside thorough clinical validation and feedback from medical professionals. This will include exploring Vision Transformers for image classification, multi-view mammogram analysis, and developing three-dimensional mammogram reconstructions to enhance diagnostic processes significantly.

## Chapter 2

### Literature Review

The concept of employing computer-based systems for medical image analysis is not novel. In the past ten years, numerous innovative techniques and advancements have been introduced in breast imaging. These tools are currently extensively utilized to aid clinicians in generating second opinions for diagnostic procedures and automatically evaluating all potential abnormalities in breast images. This chapter discusses the previous innovative research on the detection of breast cancer in medical images.

#### 2.1 Data Augmentation

Data scarcity is a critical challenge for DL researchers. Acquiring sufficient labeled data is a key challenge for training CAD models using ML or DL techniques. Considering the current health sector scenario, acquiring a label dataset is challenging. DL models are now prevalent in computer vision tasks. DL models have surpassed human physician accuracy. Such models rely on the availability of sufficient training datasets. Some of the datasets related to breast cancer are not publicly accessible, imbalanced, or insufficient for the training of DL models. Such issues make DL models unreliable. In real life, a lot of mammograms are generated daily worldwide. These mammograms can contribute to the ongoing research on breasts. However, certain obstacles limit using real mammograms for deep-learning models. The prominent barriers include patient privacy, mammogram variability, labeling and annotations of the mammograms, and cost and time constraints. Creating new training data artificially from preexisting training data is known as data augmentation. Researchers are using various data augmentation techniques to solve these problems.

Zeiser, et al. [28] have developed a model for segmenting masses in mammograms using DL techniques. This work consists of four phases. First, they removed irrelevant information and enhanced the contrast of the 7989 mammograms. The researchers utilized a benchmark dataset known as the Digital Database for Screening Mammography (DDSM). Next, they collected regions of interest. They performed

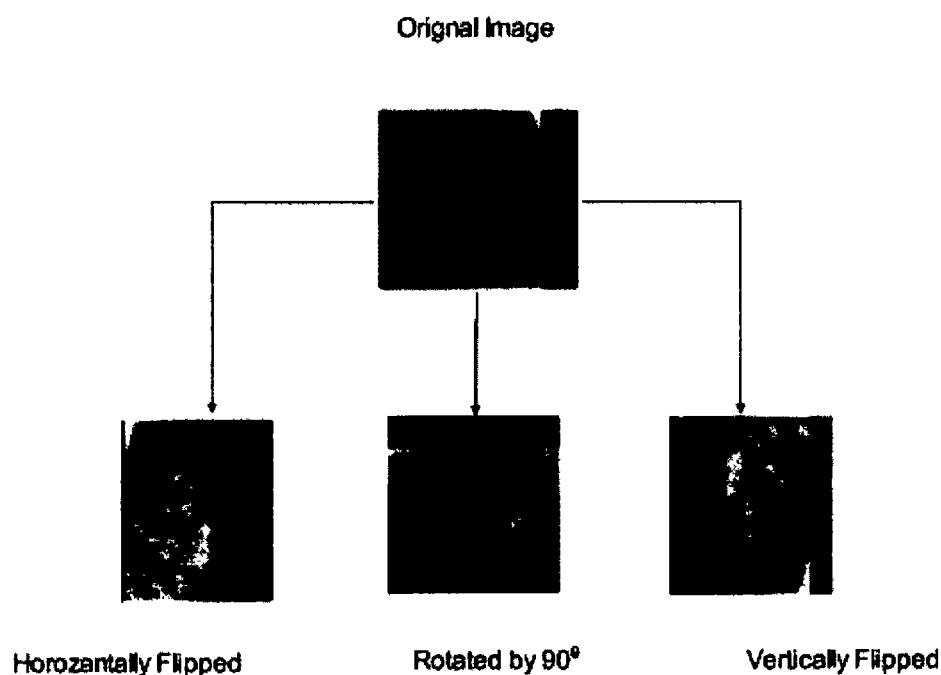
horizontal mirroring, resizing, and zooming to augment the data. They trained U-Net models. In the last phase, the model was tested and evaluated. The model achieved an Accuracy of 85.95%, a specificity of 80.47%, a sensitivity of 92.32%, and a Dice index of 79.39%. The research accurately detects lesions, their format, and their characteristics. They claim the full usage of the DDSM database. The model classifies high-intensity pixels as masses, which is not always true in the case of dense breasts. The model's performance could be improved if more data is provided for the training process. Guan *et al.* [29] utilized GAN to produce artificial mammograms using the DDSM dataset. The authors collected 2620 Regions of Interest ROIs from the original dataset. They also used the traditional data augmentation technique, affine transformation, to generate the same number of ROIs. The exact number of ROIs were generated through GAN. They compared the statistical properties of GAN-generated (ROIs) with those of ROIs of affine transformation. Specifically, they looked at these two ROIs' mean, standard deviation, skewness, and entropy. They found that the statistical properties of GAN-generated ROIs were more like those of real ROIs. With GAN ROIs, they achieved 80% accuracy with a 3.6% improvement over the affine data augmentation method. They concluded that synthetic data generated by GAN can preserve critical properties and patterns from the original data. Swiderski *et al.*[30]proposed a DL model for mammogram recognition. The model performs a special autoencoder GAN for data augmentation. The generator produces additional images in a perfect way for training the model. The final set of original and generated images are given as input to CNN for classification. A total of 11,218 ROI of mammograms from DDSM were used in the experiments. They reported an average accuracy in detecting abnormal vs healthy cases of 89.71%. Specificity stood at 80.58%, while sensitivity and AUC were 93.54% and 0.9410, respectively. The authors claim its novelty in the data augmentation compared to the existing models. However, their proposed model AGAN is learned only on normal mammograms. The model does not consider other mammographic datasets. The extensive literature study highlights that the researchers focused on the different dimensions of augmentation. Some researchers used to augment only normal classes, while others concentrated on binary class augmentation, where they worked with normal and abnormal classes. However, the available datasets typically have normal, benign, and malignant mammograms. In addition, the researchers used GAN architecture to augment the dataset.



### 2.1.1 Traditional Mammogram Augmentation Techniques

Flipping, scaling, rotation, and contrast adjustment techniques inflate the datasets. Flipping the images is useful as breast abnormalities can occur in either breast. The model learns the abnormalities by flipping the images regardless of their location within the breast mammograms. Similarly, scaling images simulate variations in breast size and help the model learn to detect abnormalities at different scales; cropping can focus on specific regions of interest, such as the breast tissue or suspicious areas, to provide more targeted information to the model. Rotating mammogram images by small angles within the range of 10-15 degrees is helpful to detect abnormalities at different orientations, as shown in Figure 2.1. Adjusting the contrast of mammogram images is also used by researchers for the detection of masses in mammograms with various levels of contrast. Zeiser, et al. [28] have developed a CAD system for segmenting masses in mammograms using DL techniques. The whole research work consists of four phases. First, they removed irrelevant information and enhanced the contrast of the 7989 mammograms of the Digital Database for Screening Mammography (DDSM) database, acquiring the region of interest. Data augmentation with horizontal mirroring, resizing, and zooming was accomplished in the second phase. They utilized six U-Net models based on their architecture. The model was assessed in the final stage. Their findings indicated an accuracy rate of 85.95%, a specificity rate of 80.47%, a sensitivity rate of 92.32%, and a Dice index value of 79.39%. The primary contribution of this study is the capacity to precisely identify the lesion, its structure, and its attributes. Compared to the other related work, they claim the total usage of the DDSM database. The model classifies high-intensity pixels as masses, which is not always true in the case of dense breasts. The model's performance could be improved if more data is provided for the training process. In this study [31], the authors evaluated the performance of two breast cancer detection models, AlexNet and Inception V3, using the CBIS-DDSM and INbreast datasets. They utilized pre-processing methods, including denoising and contrast limited adaptive histogram equalization (CLAHE). Data augmentation was used to reduce overfitting and enhance the models' generalization capability. The reported results demonstrated that using augmented data significantly increased the testing accuracy for low-quality input images, from 60.26% to 86.7%. The system was evaluated solely based on ROIs. This research [32] aims to validate a deep neural network model for breast cancer detection using local and one

benchmark dataset DDSM. The local dataset was collected from Osaka City University Hospital and Medcity21 Clinic from January 2006 to December 2017. The experiments utilized 3,179 images in the hospital development dataset, 491 images in the hospital test dataset, 2,821 images in the clinic test dataset, and 1,457 malignant images in the DDSM benchmark. The mean false positive indications per image (mFPI) for the hospital datasets were 0.45, with an accuracy of 0.86 and a specificity of 0.88. The clinical dataset's mFPI was 0.47, with an accuracy and specificity of 0.85 each. The datasets exhibited an AUC of 0.84, indicating commendable overall performance. Additionally, the sensitivity was also 0.84. The model exhibited a low rate of false positives. The clinical data obtained from the hospital was composed of normal cases.



**Figure 2.1: Illustration of Traditional Mammogram Augmentation**

Using a local dataset of grey-scale breast mammograms acquired from Gachon University Gil Medical Centre, South Korea, Kim and Kim [33] proposed a deep-learning model in their study. They used ROIs from the gathered dataset to train a model. A deep-learning model was employed to detect malignant masses in 8-bit mammogram images. The input images were preprocessed by applying a scaling parameter to adjust the intensity before training the CNN model. The results indicated that the model achieved the highest area under the receiver operating characteristic

curve, with a value of 0.897. The diversity and standardization of local datasets are always low. Research conducted by [34] used the You Only Look Once (YOLO) v3 model for breast cancer detection using the benchmark INbreast mammography dataset. The authors conducted a performance comparison between ResNet and Inception as feature extractors of YOLO. YOLO demonstrated a detection rate of 89.4% for masses in the INbreast dataset, with an average precision of 94.2%, as indicated by the results. The model also attained an accuracy of 84.6% in accurately categorizing masses as benign or malignant. The System is limited to only binary classification. Sabani *et al.* [35] created a breast cancer detection system using a specific dataset called Picture Archiving and Communication System (PACS). The dataset consisted of 32,579 mammographic images collected from 2010 to 2019. Seventy percent of the image patches were used to train a deep CNN model. The remaining 20% and 10% data were used for validation and evaluation. The model exhibited a specificity of 91.3% and a sensitivity of 84.0%. However, sensitivity was slightly lower than that of human readers, with reader 1 and reader 2 achieving a sensitivity of 88.0% and 88.0%, respectively. Table 2.1 presents the comparison of various traditional mammogram augmentation techniques in breast cancer research.

Table 2.1: Comparison of Traditional Mammogram Augmentation Techniques

| Ref. & Year              | Dataset              | Samples                       | Technique  | Limitation   |
|--------------------------|----------------------|-------------------------------|--|--|
| Zeiser, et al. [28] 2020 | DDSM                 | 7989                          | Mirroring, zooming, and resizing                     | Redundant feature learning due to flipping for data augmentation |
| Razali et al. [31], 2021 | CBIS-DDSM & INbreast | 1318 & 410                    | Rotation, flipping, and shearing                     | Limited evaluation to ROIs only                                  |
| Aly et al. [34] 2022     | INbreast             |                               | Flipping and rotation                                | Limited to only binary classification.                           |
| Ueda et al. [32] 2022    | DDSM & Local Dataset | 3179, 419, 2821, 1457         | Flipping, rotation, shifting, shearing, and scaling  | Dataset imbalance with dominant normal cases                     |
| Kim and Kim [33] 2022    | Local                | 300 images with 281 malignant | Shifting   | Low diversity and lack of standardization in local datasets      |
| Sebani et al. [35] 2022  | Local                | 32579                         | Flipping, zooming, rotating, and brightness changing | Lower sensitivity compared to human expertise                    |

### 2.1.2 GAN-based Mammogram Augmentation Techniques

The field of ML has experienced substantial progress in the last decade due to the advancements in deep neural networks and their exceptional performance in various medical imaging tasks. Meanwhile, advancing GANs have led to unprecedented generative modeling and data synthesis quality. GANs generate remarkable realistic visuals that accurately correspond to the data they were trained on.

Wu *et al.* [36] conducted a study to tackle the issue of limited and unevenly distributed data in the detection of breast cancer. They utilized the Optimam Mammography Image Database, a publicly accessible dataset from the United Kingdom (UK), for their research. The dataset comprises 8282 malignant, 1287 benign, and 16887 normal mammographic images. The data was partitioned into three sets: 60% for training, 20% for validation, and 20% for testing. A contextual GAN model was trained to enhance the dataset by incorporating a self-attention mechanism. They employed a combination of conventional and GAN-based augmentation techniques. Their model, enhanced with a GAN, achieved an AUC of 0.846. The model conducts binary classification to distinguish between normal and malignant cases. The training process involves utilizing individual images, which can result in less stable learning and increased susceptibility to atypical or deceptive images. Desai *et al.* [37] developed the DCGAN model using a benchmark DDSM dataset. They utilized 218 for training and 47 each for testing and validation. Their experiments reported an accuracy of 78.23% when the model was trained on original images; simultaneously, the combination of synthetic and original images produced an enhanced accuracy of 87% with an improvement factor of 8.77. The authors show that GAN is a workable choice for training such models with a data shortage. Alyafi *et al.* developed a DCGAN model for breast mass augmentation using a subset of 80000 images from the UK-based OPTIMAM Mammography Image Database (OMI-DB) [38]. The authors demonstrate the performance of a classifier in an imbalanced dataset with and without synthetic data in the experiments. They created breast mass patches with  $128 \times 128$  pixel dimensions using a modified version of DCGAN. GAN augmentation was compared to traditional augmentation. The results show that using DCGANs with flipping augmentation improves the F1 score by up to 0.09 compared to the original mammographic images. The job can be expanded to include other similar tasks. Their work is limited to small

mammogram patches. The study focuses on generating mass patches of  $128 \times 128$  pixels, which might not capture the full context of larger mammogram images. The smaller size could limit the detail and contextual information in the synthetic images. Shen *et al.* [39] developed a GAN-based system using a benchmark DDSM dataset and a local dataset collected from Nanfang Hospital, China. The study aimed to address the issues of limited data in medical image analysis. They designed a model to generate labeled images based on contextual information within the breast mammograms. The model was evaluated, and the results showed that their augmentation technique increased the diversity of the dataset and achieved an improvement of 5.03% in the detection rate. A model is a viable option for generating labeled breast images. However, the limited number of 765 mammograms with malignant mass findings from the DDSM dataset may not encompass the complete spectrum of variability observed in actual clinical environments. They used traditional data augmentation. The private dataset is sourced exclusively from Chinese patients, which may introduce bias in the model. A mammogram recognition model based on DL was presented in [30]. The model utilizes an autoencoder-generative adversarial network (AGAN) specifically designed for data augmentation. The generator efficiently generates supplementary images to enhance the training of the model. CNN is fed with the last batch of original and generated images for classification. The experiments utilized 11,218 ROI (regions of interest) from mammograms in the DDSM dataset. Their average accuracy in distinguishing abnormal from healthy cases was 89.71%. The specificity was measured at 80.58%, while the sensitivity and AUC were 93.54% and 0.9410, respectively. The primary significance of the work lies in its originality regarding data augmentation, distinguishing it from other deep-learning approaches. The AGAN model is exclusively trained using normal data. The model does not consider additional mammographic datasets. Another study generated breast mammograms with GANs [40]. The prime purpose was to detect mammographically occult (MO) breast cancer in women with dense breasts. The researchers employed CNN. The network was trained on processed mammographic images from Radon Cumulative Distribution Transform (RCDT) 1366 processed mammograms collected from the University of Pittsburgh Medical Center, United State of America (USA). They reported an AUC of 0.77. The system can identify patients for further screening in the early detection of MO-related cancer. However, they did not consider benchmark datasets. The authors developed a StyleGAN 2 system using 105,948 normal mammograms collected from Asan Medical

Center, Korea, from January 2008 to December 2017[41]. The GAN-generated images were assessed using the Fréchet Inception Distance (FID) metric, which yielded a value of 4.383. Additionally, the Inception Score was determined to be 16.67. The Multi Scale Structural Similarity Index MS-SSIM was 0.39, while the average Peak Signal-to-Noise Ratio (PSNR) was 31.35. Their model has achieved a satisfactory level of accuracy in reproducing real images. The system was only limited to normal mammographic local images. These studies are summarized in Table 2.2.

Table 2.2: Comparison of GAN-based Mammogram Augmentation Techniques

| Ref. & Year                       | Dataset                                     | Samples | Technique      | Limitation  |
|-----------------------------------|---|---------|----------------|---|
| Wu <i>et al.</i> [36] 2020        | Optimam Mammography Image Database          | 26456   | Contextual GAN | Unstable learning with single image Training.               |
| Desai <i>et al.</i> [37] 2020     | DDSM  | 287     | DCGAN          | Limited evaluation of two batch sizes.                      |
| Alyafi <i>et al.</i> [38] 2020    | OPTIMAM Mammography Image Database (OMI-DB) | 80000   | DCGAN          | Limited context from small image patches.                   |
| Shen <i>et al.</i> [39] 2021      | DDSM and Local dataset                      | 11218   | GAN            | Limited dataset augmentation                                |
| Swiderski <i>et al.</i> [30] 2021 | DDSM  | 11218   | AGAN           | Misclassification risk in GAN with autoencoder architecture |
| Lee <i>et al.</i> [40] 2022       | Local Dataset (USA)                         | 1366    | CGAN           | They did not consider benchmark datasets                    |
| Park <i>et al.</i> [41]. 2023     | Local dataset (Korea)                       | 105,948 | StyleGAN 2     | The system was only limited to normal mammographic images.  |

2.2 Synthetic Data Validation

Training the datasets triggers the fundamental problem of overfitting when a data augmentation mechanism is used. This is done for reasons that involve taking more data from original datasets using augmentation. Augmentation techniques synthetically inflate the required training dataset size. This process is accomplished either through oversampling or data-wrapping techniques. Augmentation may be performed by color or geometric transformation, erasing, and adversarial training. The oversampling techniques generate synthetic samples and are mixed with original training data. This process may combine different images, augment the feature space, or go through GANs. Data wrapping and oversampling-based augmentation do not lead to a mutually

exclusive dichotomy. In some cases, GAN-generated data can be stacked with a random cropping mechanism to add synthetic data [37]. PM Shah *et al.* [42] used an unsupervised data technique to empirically annotate and validate the synthetic x-rays (of Corona Virus Disease of 2019 (COVID-19) generated by GANs. All the synthetically generated images are shuffled with the original images to mix up the original and synthetic images. The combined dataset is then split into three clusters (classes) using k-means clustering. The resultant dataset presented a more generic form. The research in [43] focuses on assessing GAN-synthesized periapical images for identifying C-shaped root canals. Utilizing StyleGAN2-ADA, it aimed to overcome the limitations of small training datasets in dental imaging. Image quality was measured through FID scores and a visual Turing test, indicating the images' realism. The study found that the neural network's diagnostic performance improved when GAN data supplemented real images. This suggests GANs are a viable tool for enhancing data availability in dental imaging, particularly for complex cases like C-shaped canals. Despite the popularity of the GAN in generating synthetic imaging data, the researchers are only relying on the discriminator part of the GAN. The discriminator is a function that validates the authenticity of synthetic images. However, depending on the synthetically generated dataset is debatable.

### 2.2.1 Visual Inspection

Visual inspection is one of the fundamental methods for validating synthetic data generated by DL models, particularly GANs. This technique involves the manual evaluation of synthetic images by human experts to assess their quality and realism[44]. Despite being somewhat subjective, visual inspection remains a critical initial step in the validation process due to its straightforward application and immediate feedback. Visual inspection provides a direct means of assessing the visual quality of synthetic data. It allows researchers to quickly identify obvious flaws or artifacts that automated metrics might overlook. This method is particularly valuable in the context of medical imaging, where the visual fidelity of synthetic images is paramount for their potential use in clinical applications[45].

In the domain of breast cancer detection, several studies have utilized visual inspection to validate synthetic mammograms generated by GANs. For instance, Korkinof *et al.* [46] employed expert radiologists to evaluate the visual quality of synthetic

mammograms, finding that the images generated were indistinguishable from real ones to a significant extent. This validation step was crucial in establishing the credibility of their synthetic data for augmenting training datasets in mammogram classification tasks.

Similarly, Han et al. [47] used visual inspection as part of their validation framework for synthetic mammograms. Expert radiologists assessed the synthetic images for typical features of breast tissue and potential abnormalities. The study reported high levels of agreement among the radiologists regarding the realism of synthetic images, reinforcing the utility of visual inspection in the validation process.

While visual inspection is valuable, it has inherent limitations. The subjective nature of this method means that results can vary depending on the expertise and experience of the evaluators. Moreover, visual inspection is labor-intensive and not scalable for large datasets. To mitigate these issues, it is often used in conjunction with other validation techniques, such as statistical analysis and feature distribution comparisons.

To enhance the robustness of synthetic data validation, visual inspection is frequently combined with other methods. For example, in the study by Xu et al, [48] visual inspection was complemented by quantitative measures such as the SSIM and PSNR. This multi-faceted approach provided a more comprehensive validation of the synthetic data, ensuring both visual and statistical fidelity.

Visual inspection remains a crucial component of the synthetic data validation toolkit, particularly in fields requiring high visual fidelity like medical imaging. Its application in breast cancer detection has demonstrated its effectiveness, despite its subjective nature. Future research should continue to refine this method, potentially integrating it with automated techniques to improve scalability and consistency.

### **2.2.2 Statistical Analysis**

Statistical analysis is a crucial method for validating synthetic data generated by DL models, such as GANs. This technique involves using statistical measures to compare the properties of synthetic data with those of real data, ensuring that the synthetic data accurately represents the underlying distributions and characteristics of the real



dataset[49]. Statistical methods provide an objective way to quantify the similarity between real and synthetic data. Commonly used metrics include mean and variance comparisons, hypothesis testing, and more advanced techniques like SSIM and PSNR. These methods are essential for validating the quality and utility of synthetic data, particularly in fields where data accuracy and integrity are critical, such as medical imaging. In the context of breast cancer detection, several studies have employed statistical analysis to validate synthetic mammograms generated by GANs. For instance, Cha et al. [50] used statistical measures to compare the pixel intensity distributions of real and synthetic mammograms. By calculating the mean and variance of pixel intensities, they demonstrated that the synthetic images closely matched the statistical properties of the real images. This statistical validation was a key step in confirming the synthetic data's quality and suitability for training machine learning models.

Another study by Furtney et al. [51] utilized more sophisticated statistical techniques to validate synthetic mammograms. They employed SSIM and PSNR metrics to compare the structural and visual quality of synthetic images against real mammograms. The SSIM index measures the similarity between two images by considering changes in structural information, luminance, and contrast. The PSNR metric quantifies the difference between the images by measuring the peak error. Their results showed high SSIM and PSNR values, indicating that the synthetic images were highly similar to the real ones in both structure and quality.

Furthermore, statistical analysis can be extended to feature distribution comparisons, where the distributions of key features (e.g., texture, shape) are compared between real and synthetic data. For example, Shivhare et al. [52] performed histogram analysis of texture features in real and synthetic mammograms. By comparing the histograms, they verified that the synthetic data preserved the essential textural characteristics of the real data. This method provided additional assurance that the synthetic data could effectively augment real datasets for training purposes.

Statistical analysis is not without its challenges. The choice of appropriate statistical measures and the interpretation of results require careful consideration. Additionally, while statistical similarity is important, it does not always guarantee the clinical relevance of synthetic data. Therefore, statistical analysis is often used in conjunction

with other validation methods, such as visual inspection and performance evaluation, to provide a comprehensive validation framework.

Statistical analysis plays a pivotal role in the validation of synthetic data generated by GANs for breast cancer detection. By quantifying the similarity between real and synthetic data through various statistical measures, researchers can ensure the synthetic data's quality and reliability. The combination of statistical analysis with other validation methods provides a robust approach to validating synthetic data for use in medical imaging applications.

## **2.3 Issues with Synthetic Data**

Following the discussion on data generation and validation techniques, it is important to address the challenges that come with using synthetic data, particularly in medical imaging and breast cancer detection. While synthetic data generation can alleviate issues like data scarcity and privacy concerns, it introduces its own set of challenges [53].

### **2.3.1 Lack of Realism**

Synthetic data generation techniques, such as GANs, are widely used to create artificial mammograms. However, despite technological advancements, synthetic data often lacks the complexity and realism of real-world data. In breast cancer imaging, small differences in mammographic data can have a significant impact on the diagnostic process. The synthetic datasets may fail to capture subtle variations in tissue, limiting their utility in creating robust models. As a result, models trained on synthetic data may perform well in controlled settings but struggle to generalize to real-world cases.

### **2.3.2 Validation Challenges**

Another key issue is the validation of synthetic data. While visual inspection and expert review can help confirm the realism of the images, the lack of comprehensive annotations remains a significant barrier. In many cases, annotations are either unavailable or expensive to create, which makes it difficult to fully validate the synthetic datasets. Without these annotations, models trained on synthetic data might

not learn the subtle distinctions required for accurate classification, thus impacting their reliability in real-world applications.

### **2.3.3 Bias and Overfitting**

Synthetic data can also introduce bias into machine learning models if it does not adequately represent the diversity found in real mammographic datasets. Models trained on such data are at risk of overfitting, performing well on the synthetic images but poorly on actual clinical data. In the context of breast cancer detection, this is particularly concerning, as it could lead to misdiagnoses or missed cases. Ensuring that synthetic datasets are diverse enough to capture a wide range of real-world scenarios is essential to mitigate these risks.

### **3.2.4 Quality Control**

Maintaining the quality of synthetic data throughout the generation process is another critical issue. Poor quality control can lead to synthetic mammograms that do not accurately reflect the characteristics of real images. This can directly affect the diagnostic accuracy of models trained on these datasets. Quality control is especially important in medical imaging, where small inaccuracies in the data can have significant consequences for patient care.

## **2.4 Ensemble Classification of Images for Cancer Detection**

Classification is a widely recognized method for categorizing samples into distinct groups using information from a training dataset. It entails making forecasts regarding the category of the sample. In mammography, classification is employed to forecast the type of mass by analyzing the extracted characteristics from the image.

A study [54] proposed a model based on a multi-context ensemble CNN. The models were separately trained on image patches having dimensions  $96 \times 96$ ,  $48 \times 48$ ,  $24 \times 24$ , and  $12 \times 12$ . These patches are then combined. The ensemble architecture can find and locate abnormalities. Experiments were performed on INbreast and E-ophtha datasets. Microcalcification in mammograms from the INbreast data set was detected by the model with better results. The results were validated through 2-fold cross-validation. The proposed system achieved a mean sensitivity of 83.54% for INbreast and 81.62%

for E-ophta. The research findings indicated a significant improvement in the accuracy of detecting small masses compared to a single CNN. The model detects only microcalcification. Tan and Le [55], in their work, have conducted a systematic analysis of model scaling. They found that optimizing the network depth, width, and resolution can lead to significant performance improvements in the model. They proposed a highly effective compound but simple scaling technique. This compound method allows the model to focus on more relevant regions with more object details. They compared the ImageNet performance of different scaling methods for the baseline EfficientNet-B0 network. The authors stated that other scaling methods improve the accuracy at the cost of more FLOPS. Their approach showed 2.5% improved accuracy compared to another single-dimensional scaling. The main success factor of their model was its ability to focus on more relevant regions and object details as compared to similar models, which were either due to lack or inability to get such details.

Li, et al. [56] explained that the primary attention mechanism in CNNs aims to pinpoint the most important features and avoid redundancy. The research incorporated attention mechanisms in CNNs by utilizing corresponding attention maps. The researchers introduced a novel called AG-CNN attention-based CNN to detect Glaucoma. In addition, they created a comprehensive Glaucoma database called the Large-Scale Attention-Based Glaucoma (LAG) database. This database contains a total of 5824 fundus images, with 2392 being positively identified as Glaucoma and 3432 not having Glaucoma. Their experiments demonstrated that attention maps successfully identified pathological areas, enhancing accuracy, sensitivity, specificity, and F2-score by 4.5%, 4.3%, 4.7%, and 4.7%, respectively. The proposed model exhibited superior performance compared to other innovative models, and creating the LAG dataset, complete with labels and attention maps, represents a noteworthy contribution. Nevertheless, the model's efficacy declined when evaluated on the Retinal Image Database for Optic Nerve Evaluation (RIM-ONE) database.

Khamparia, et al. [57] used a transfer learning process for detecting breast cancer. The researchers conducted experiments utilizing 2D mammographic images and introduced a novel model named the modified Visual Geometry Group (MVGG). A total of 10,713 patches were obtained from DDSM, consisting of 4,506 pathological patches and 6,027 nonpathological patches. The data was partitioned into three sets: 75% for training, 10%

for validation, and 15% for model testing. Before training, they conducted data augmentation through geometric transformations. Subsequently, the hybridization of MVGG and ImageNet took place. The hybrid model attained an accuracy rate of 88.3%. The hybrid pre-trained network that was suggested demonstrated superior performance compared to existing CNNs. The model exclusively conducts binary classification.

The researchers in [58] developed a model to reduce false positives and negatives during mammogram screening for breast cancer detection. They have trained the Visual Geometry Group of the University of Oxford (VGG) 16 models on Mammographic Image Analysis Society (MIAS) datasets. A total of 322 mammograms are classified into normal and abnormal. They performed data augmentation using traditional flipping, rotation, and resizing techniques. A total of 2600 images were used in the experiments. The model was trained on 70% of the data, while 15% each was used for testing and validation. Apart from VGG16, EfficientNet, GoogleNet, and AlexNet were also used in the experiments. VGG16 reported the highest accuracy of 75.46%, followed by EfficientNet, with an accuracy level of 72.29%. The model's technique for classification has the potential to assist in the timely detection of breast cancer-related diseases.

The study presented in [59] outlines a method for the ISIC 2019 Skin Lesion Classification Challenge. Their first task focused on skin lesion classification based on dermoscopic images. Their second task incorporated additional patient metadata along with dermoscopic images. The training dataset included 25331 dermoscopic images collected from multiple sources, followed by various preprocessing techniques. The researchers employed an ensemble DL model that utilized models. Models like EfficientNet, SENet, and ResNet were used. The largest EfficientNet produced the best results for skin lesion classification in an ensemble manner. The final ensemble automatically chose the SENet and ResNet models, indicating that incorporating architectural variability is advantageous. The average sensitivity reported for task 1 was  $71.7 \pm 1.7$ , while for task 2, it was  $73.4 \pm 1.6$ . Nevertheless, the models tended to overfit due to the problem of incomplete data for the unidentified class, thereby impacting the efficacy of their model.

In [60], researchers have used ensemble learning methods to enhance the accuracy of their model by combining multiple EfficientNets. They utilized the DDSM dataset.

First, they performed data preprocessing followed by transfer learning. For pre-training, they used an ensemble model of three EfficientNets. Each of these models received an equal portion of the Common Objects in Context (COCO) dataset. One-third of the dataset was allocated for training for each model. To guarantee that every training dataset was distinct, the set was divided into three parts for the DDSM part training, and each pre-training model was trained in two parts. Finally, the non-maximum suppression (NMS) method was used to integrate the model output. They obtained a recall rate of 91.8% and an accuracy rate of 87.6%, suggesting that the integrated architecture can successfully raise the detection model's accuracy level. The researchers trained the model on two datasets, but the processing speed was sluggish.

In [61], Devi Vijayan presented an ensemble learning for the classification of masses. Their work utilizes three distinct feature extraction methods, specifically scale-invariant feature transform (SIFT), local binary pattern (LBP), and grey-level co-occurrence matrix (GLCM), to extract features from the regions of interest (ROIs). In their experiments, four different classification models, including artificial neural network (ANN), support vector machine (SVM), extreme gradient boosting machine (XG Boost), and random forest (RF), are used for each feature descriptor. They proposed a technique that tests 12 combinations. The ensemble framework is constructed separately for each BI-RADS density category. They performed parameter tuning using grid search to prevent overfitting of the model. In this research, a total of 48 models were evaluated for BI-RADS density classification. The proposed approach achieved an overall average accuracy of 69.2%.

Yu et al. [62] employed deep neural network (DNN) models trained to identify breast lesions through image analysis. The diagnostic performance of the Benchmark Breast Cancer Digital Repository (BCDR) was assessed using 406 lesions. The accuracy of the predictions and the average area under the curve were used as measures. According to the experimental findings, GoogleNet achieves the highest performance, with an accuracy rate of 81% and an AUC value of 88%. AlexNet ranks second, achieving an accuracy of 79% and an AUC of 83%. CNN3 ranks third, attaining an accuracy of 73% and an AUC of 82%.

According to the BI-RADS classification method, Boumaraf et al. [63] presented a new CAD for classifying breast masses in mammograms. The project aims to enhance

feature selection by utilizing a modified genetic algorithm to enhance the CAD system's accuracy. The examination utilizes 500 mammogram images sourced from the benchmark DDSM dataset. The study reported the following metrics: classification accuracy of 84.5%, positive predictive value of 84.4%, negative predictive value of 94.8%, and Matthew's correlation coefficient of 79.3%.

Farhan et al. [64] conducted research on benchmark dataset Mini-MIAS for feature extraction methods like LBP and Histogram of Oriented Gradient (HOG) and Gray-Level Co-Occurrence Matrix (GLCM) to analyze bulk tissue and ROI characteristics. Classifiers with the best results used these features with the highest accuracy (92.5%) using the LBP approach and logistic regression classifier at ROI (30x30). The SVM classifier yielded the best results with the HOG technique, with 90% accuracy at ROI (30x30). The K-Nearest Neighbors (KNN) classifier delivers the most significant impacts with GLCM, achieving 89.3% accuracy at ROI (30x30).

Song et al. [65] created a DL-based mixed feature CAD approach to classify mammographic masses as normal, benign, or cancerous. Deep Convolution Neural Network (DCNN) feature extractors graded three breast masses. The SVM classification techniques performed worse than the Extreme Gradient Boosting (XGBoost) classifier. They conducted experiments on 11562 ROIs from the DDSM dataset. Thus, when XGBoost was employed as a classifier, the overall accuracy was identified as 92.80%, and for malignant tumors, it was 84%, with reasonable and best results. These results suggest that the proposed strategy may help diagnose instances that are hard to classify on mammographic images.

Arora et al. [66] utilized a machine-learning neural network algorithm to identify benign and malignant tumors in CBIS-DDSM autonomously. The network achieved an accuracy of 0.88 and an AUC of 0.88 after being trained and evaluated to distinguish between benign and malignant tumors. The results indicate that the proposed breast cancer classification system is resilient and promising.

Zhang et al. [67] assessed the efficacy and resilience of their model by employing CBIS-DDSM and INbreast, two widely recognized mammographic datasets. The AdaBoost algorithm was employed to identify and classify benign and malignant breast masses. The CBIS-DDSM model demonstrated an accuracy of 90.91% and a sensitivity

of 82.96%. Their model achieved an accuracy of 87.93% on the INbreast dataset, with a sensitivity of 57.20%.

Patil and Biradar [68] used a deep-learning model for breast cancer detection. They used CNN with Recurrent Neural Network (RNN) to classify masses into normal, benign, and malignant. They reported an accuracy of 90% and a sensitivity of 92%. The precision and F1 scores were 78% and 84%, respectively.

A detailed multi-fractal analysis study extracts texture information in mammography images for BI-RADS density classification [69]. The suggested approach is tested on 409 mammography images from the INBreast benchmark dataset. Experimental results reveal that multi-fractal analysis and LBP-based cascaded feature descriptors have higher classification accuracy than individual feature sets. The study reported an accuracy of 84.6 % with an AUC of 95.3%. Although the existing techniques showed ample accuracy, there is still room for improvement, as the accuracy of these models is still under 90%. In critical health issues, accuracy plays a vital role. In addition, the research's primary focus should be improving the false positives.

In [70], the authors used an ensemble approach by combining DenseNet, EfficienetNet, and XceptionNet. The ensemble approach demonstrated superior performance compared to individual models, achieving an accuracy of 88.33%, an F1 score of 75.88, a recall of 76.29%, and a precision of 85.62%. This approach resulted in an overall performance improvement of 5% compared to the individual models.

X Wang et al. [94] devised an innovative deep-learning model that surpasses current approaches in accurately detecting breast cancer from slide images. The model exhibited exceptional performance, with an accuracy of 86.21%, precision of 85.50%, and sensitivity of 85.60%. Compared to other ML and DL methods, it also attained a specificity of 84.71%, an F1-score of 88%, and an AUC of 0.89, indicating its strength and dependability in minimizing diagnostic mistakes.

Ammar et al. [71] explore a DL approach to identify breast cancer from biopsy microscopy images, utilizing deep convolutional networks. The study investigates the impact of data pre-processing techniques on model performance and introduces an ensemble method to enhance accuracy. Experimental findings reveal that models such as Densenet169, Resnet50, and Resnet101 show significant improvements in accuracy



when coupled with data augmentation and segmentation techniques. The ensemble approach further enhances model performance, achieving a peak accuracy of 92.5%.

Razali et al. [66] presented a study that innovates computer-assisted mammography by integrating deep ensemble transfer learning with a neural network classifier for feature extraction and classification. This approach, which preprocesses mammogram images before using them in an ensemble model, demonstrates significant efficacy with precision and an AUC of 88%, suggesting a promising direction for breast cancer diagnostic tools.

The study [72] examines a stacked ensemble method in CADx systems for the detection of breast cancer to improve mammographic image classification. The technique combines multiple classifiers to improve accuracy and reduce errors. The results show that the ensemble approach outperforms single classifiers, achieving 85% accuracy, 86.67% sensitivity, 83.87% precision, and 83.33% specificity, indicating its potential efficacy in mammogram classification in CADx systems.

Nakach et al. [73] explore transfer and ensemble learning to classify breast cancer histological images in the BreakHis data set at different magnifications. It employs bagging ensembles combining pre-trained DL techniques with ML classifiers Multilayer Perceptron (MLP), SVM, KNN) as base learners. The research compares various configurations of these ensembles, determining the most effective one through the Scott-Knott statistical test and Borda Count voting. The optimal ensemble, which features a combination of DenseNet201, 200× magnification, and MLP, achieved a mean precision of 93.98%, underscoring the potential of hybrid DL approaches to classify histopathological images of breast cancer.

In another study, Hekal et al. [74] introduced an ensemble deep-learning system designed for the early detection of breast cancer. The system, which combines four transfer learning CNNs with a binary support vector machine (SVM), offers improved detection of small nodules. The effectiveness of the set is demonstrated with 94% and 95% precision in distinguishing between malignant and benign classes and nodules, respectively, using the CBIS-DDSM medical image dataset. This approach shows significant improvements over existing methods in terms of accuracy. The findings of these studies are outlined in Table 2.3

Table 2.3: Classification Techniques for Cancer Detection in Breast Images

| Ref. | Dataset             | Samples        | Classifier  | Results  | Weakness                                   |
|------|---------------------|----------------|---|--|--|
| [54] | INbreast & E-ophtha | 115& 233+148   | CNN   | Sensitivity: 83.54% (INbreast), 81.62% (E-ophtha)            | Microcalcification detection limitation    |
| [57] | DDSM                | 10,713 patches | Hybrid MVGG and ImageNet                                | Accuracy: 88.3%  | Limited to patch-based data                |
| [58] | MIAS                | 2600           | VGG16, EfficientNet, GoogleNet, and AlexNet             | Accuracy: 75.46%   | Performance measured only by accuracy.     |
| [59] | Local Dataset       | 25331          | Ensemble Classifier EfficientNet, SENet, and ResNet WSL | Task1 Accuracy: 71.7 ± 1.7%, Task2 Accuracy: 73.4 ± 1.6%     | Overfitting due to missing data.           |
| [60] | DDSM                | 5,316          | Ensemble Classifier                                     | Accuracy: 87.6%, Recall: 91.8%                               | Slow processing speed.                     |
| [61] | DDSM                | 1657           | Ensemble Classifier ANN, SVM, XG Boost and RF           | Accuracy: 69.2%  | High computational complexity.             |
| [70] | DDSM                | 2620           | Ensemble DenseNet, EfficientNet, XceptionNet            | Accuracy: 88.33%, F1: 75.88, Recall: 76.29, Precision: 85.62 | Potential for further model refinement.    |
| [94] | PCam                | 277,524        | CNN and Gated Recurrent Unit (GRU)                      | Accuracy: 86.21%, Precision: 85.50%, F1: 88%, AUC: 0.89      | Need for additional classification models. |

The above techniques often face challenges such as overfitting, limited patch-level analysis, and high computational demands. Future developments could focus on integrating advanced DL architectures, utilizing larger and more diverse datasets, optimizing real-time processing, incorporating multi-modal data, enhancing explainability, and implementing continuous learning to improve detection rates and clinical applicability in medical settings.

2.5 Visualizing Lesions in Breast Imaging

Different authors have defined visualization of the model as its ability to give a quantitative understanding of the network. No well-established method is available to explain the mechanism of a DNN. Through different heatmaps, it has been revealed

that the main focus of the model is on the expected dense regions of the image. In [75], the authors have assessed trust in a developed network. This trust development is significant for applying such models in the medical field. It will pave the way to use DNN's explainability in medical diagnosis. An investigation into the development of a semi-supervised Grey-Box model on a self-training framework was conducted in [76]. This study's primary objective was to create an accurate and intelligible model. The model underwent evaluation using multiple real-world datasets from finance, medicine, and education. The researchers conducted a performance comparison of Black-box, White-box, and Grey-box models using five different labeled dataset ratios (50, 40, 30, 20, and 10%) across all six available datasets. The experimental results have shown that the proposed model is productive and operates like a Black-Box model. Additionally, they performed far better than White-Box single models. Moreover, the suggested model maintained its interpretability, akin to the White-Box model. The DNN results do not identify the specific area of the abnormal mammogram; they only indicate that the mammogram is abnormal. The radiologist will thus have difficulty identifying the cancerous cell. The current methods are limited to enhancing the model's accuracy.

## Chapter 3

### Synthetic Data Generation and Validation

With the advancement in breast imaging, the significance of robust and extensive datasets cannot be overstated. However, the challenge of acquiring sufficient real-world data has led to the emergence of innovative techniques for data augmentation. The researchers are using various data augmentation techniques to inflate existing mammographic datasets. In this chapter, synthetic data that simulate real-world scenarios is generated by DCGANs. An essential step in guaranteeing the effectiveness and dependability of breast cancer detection models is validating such synthetic data.

#### 3.1 Introduction to Mammogram Generation and Validation

Breast cancer is a common type of cancer that affects women and starts in the breast cells. It is caused by the growth of the abnormal cell in the breast, which multiplies erratically and eventually develops into a tumor. Early detection is essential because it enables prompt treatment, saving many lives [77].

In the year 2020, a total of 2.26 million new instances of breast cancer were identified globally, accounting for approximately 11.7% of the overall incidence of cancer. Currently, this form of cancer has become the most prevalent, surpassing lung cancer with a prevalence rate of 11.4%. Certain regions, namely Australia, Europe (North, West, South), and North America appear to be more significantly impacted than others[78]. Moreover, as per the reports of the World Health Organization WHO [79], the cases of cancer in women are projected to rise from 2.26 million new cases in 2020 to 3.19 million in 2040, reflecting a notable rising trend of 41% over the following two decades. Cancer is becoming more prevalent in Pakistan, with 19 million new cancers of all types recorded in 2020.

In the context of Pakistan, the annual incidence of breast cancer is predicted to exceed 83,000 cases. Annually, over 40,000 women succumb to this debilitating ailment [80].

Medical professionals use a variety of modalities, such as magnetic resonance imaging (MRI), ultrasound, and mammography, to predict breast cancer early. Mammography

of the breast is the most widely used method of breast cancer screening. It is a low-dose x-ray of the breast [81]. However, accurately interpreting mammograms is challenging as these images are complex, and the detection of abnormalities can be difficult. Human reading of mammograms can sometimes produce false-positive results or false-negative results. Variability in expertise can also lead to inconsistencies in diagnosis and treatment decisions [82].

The AI community is trying to help with breast cancer detection. Training algorithms on a large mammogram and clinical datasets conducts extensive research. These models can recognize breast cancer patterns and accurately analyze mammograms, reducing false positives and negatives and helping radiologists make better diagnoses and treatment decisions. ML models can standardize analysis and reduce radiologists' expertise [83]. Early detection of BC can be facilitated by ML algorithms, which can recognize subtle patterns and features that may not be immediately apparent to human observers. Although ML has limitations, it can detect breast cancer.

Data for training is essential to these models. The model's performance might be restricted if there are no uncommon or atypical breast cancer cases in the training data. Extensive and representative datasets are necessary for detecting cancer in various populations. If the new data is very different from the training data, these models might not be able to generalize [84]. Researchers [85] increasingly acknowledge the significance of tackling the issue of limited data representation to enhance the performance and generalizability of these models. The technique of inflating existing datasets is widely employed. Data augmentation is a procedure that involves generating additional training samples by applying different transformations to the existing data. Mammogram image augmentation encompasses techniques such as rotation, flipping, scaling, and noise. These methods aim to enhance the variability and diversity of the training data. This attribute enables the model to acquire knowledge from a broader spectrum of examples. Nevertheless, there are also disadvantages associated with this approach.

### **3.1.1 The Imperative of Data Augmentation in Medical Imaging**

The advancement of DL in medical imaging is dependent on large and diverse datasets. However, obtaining such datasets is challenging due to privacy concerns worldwide

and high labeling costs. Augmenting data addresses this by expanding the available data for training without needing new samples. It ranges from simple techniques like cropping and flipping to more sophisticated generative models, all vital for effectively regularizing DL models in medical imaging [86].

The inherent limitations of medical imaging data, such as imbalance and scarcity, must be overcome with the help of data augmentation. DL models are trained in various cases when plausible data samples are generated, especially for underrepresented classes. Diversity is essential to improve clinical tasks like classification, segmentation, and lesion detection and, consequently, diagnostic accuracy and reliability in medical imaging.

### **3.1.2 Data Scarcity in Breast Cancer Imaging**

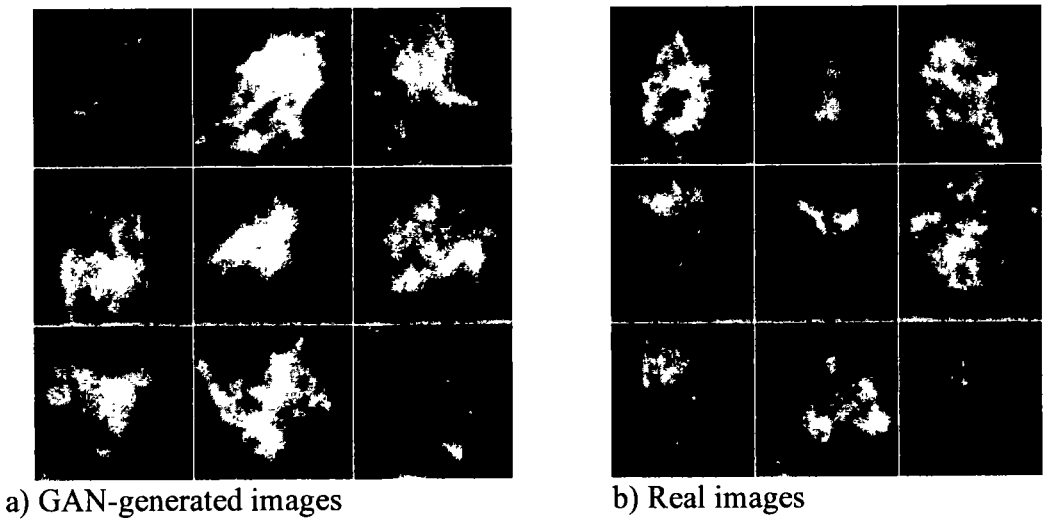
There are significant challenges associated with data acquisition in the domain of breast cancer imaging. The primary issue stems from the complexity and variety of breast cancer, which necessitates a diverse range of imaging modalities for accurate detection and diagnosis. While mammography remains the gold standard, newer technologies like digital breast tomosynthesis and various forms of MRI are being integrated, each bringing challenges in terms of feasibility and potential harm[87]. One of the critical impacts of limited data in breast cancer imaging is the potential decrease in diagnostic accuracy. With a constrained dataset, ML models, particularly those relying on DL, may not adequately learn the variability and complexity of breast cancer presentations. This limitation can lead to reduced performance in real-world scenarios, where the diversity of cases is far greater than what is represented in the training data.

### **3.1.3 Data Augmentation for Model Robustness and Generalization**

Enhancing data is crucial for enhancing the performance of ML models. Data augmentation is advantageous because it increases the variability of the data. This expansion is especially important in medical imaging, where obtaining a diverse and comprehensive dataset can be challenging due to the requirement for expert annotations, privacy issues, and the scarcity of specific conditions. Models can learn from various image representations thanks to data augmentation techniques like cropping, zooming, shearing, and adding noise [88]. These transformations decrease

the probability of overfitting by promoting the learning of invariance towards input data modifications that should not affect the output. Within DL, overfitting is a prevalent problem characterized by a model's strong performance on training data but its poor performance on unseen data. By augmenting the data, models are exposed to broader scenarios, leading to better generalization when encountering new and varied data. Additionally, there has been a trend in using generative models, like GANs, for synthetic data generation. GANs can produce realistic and varied data samples because they comprise a discriminator network trained adversarially and a generator network. This capability is particularly valuable in medical imaging, where generating new data similar to training data enhances the dataset without compromising patient privacy or requiring extensive data collection.

This study aims to use a DCGAN for synthetic data generation and confirm its effectiveness. DCGANs can create synthetic mammograms from minimal real data, as given in the below Figure 3.1



**Figure 3.1: a) GAN-generated images, b) Real images**

The primary motivation is to overcome data shortages, enhance diagnostic precision, and advance breast cancer imaging using generative models to synthesize realistic mammograms. Synthetic data could improve DL models and aid in providing patients with better care. The study aspires to improve patients' lives worldwide by improving such models' diagnostic accuracy through more robust data augmentation. The accuracy of the models helps clinicians detect breast cancer earlier.

The study's objectives include utilizing DCGANs to create synthetic mammograms to fill the data gap and increase the dataset's diversity and clinical applicability. The accuracy of DL models in detecting breast cancer can be improved by augmenting the training process with a larger and more diverse dataset. Increasing the number of cases available for comparison and validation through synthetic images will boost radiologists' confidence in their diagnostic abilities. To overcome the problem of limited annotated mammography data, the study introduces a novel methodology that uses DCGANs. It describes how to create synthetic mammograms that closely resemble the patterns found in real data. The following are the major contributions of this study:

The research extensively tests the effectiveness of the DCGAN-based synthesis in accurately replicating various mammographic features, including diverse tissue types, lesion characteristics, and breast views. The quality and authenticity of these synthetic images are carefully evaluated using mean similarity measures and standard deviation analysis, ensuring a rigorous assessment of their realism.

The study employs a systematic approach to enhance data precision by identifying and removing outliers. This is achieved through a threshold-based outlier removal mechanism, significantly bolstering the synthetic dataset's reliability. The refined dataset demonstrates clinical relevance, as evidenced by its consistency across different classes.

The reliability of the model is further corroborated through visual validation conducted by expert radiologists. Their professional assessment confirms the clinical accuracy and utility of synthetic mammograms. The study highlights the consistency of the synthetic dataset through detailed visualizations of class clustering. These visualizations demonstrate the consistency between the generated mammograms and the actual data distribution. The substantial number of images from each class passing the similarity assessment underscores the success of the validation mechanism.

### **3.2 Methodology for Synthetic Data Generation and Validation**

A comprehensive methodology used to determine the reliability of the dataset and ensure the validity of the generated mammogram classes is discussed in this section. The method includes creating mammogram classes with a DCGAN, determining



homogeneity, and removing outliers using a three-fold standard deviation threshold. The overall methodology is depicted in Figure 3.2. The systematic evaluation of the generated classes' quality and the improvement of the dataset's robustness are the goals of this study.

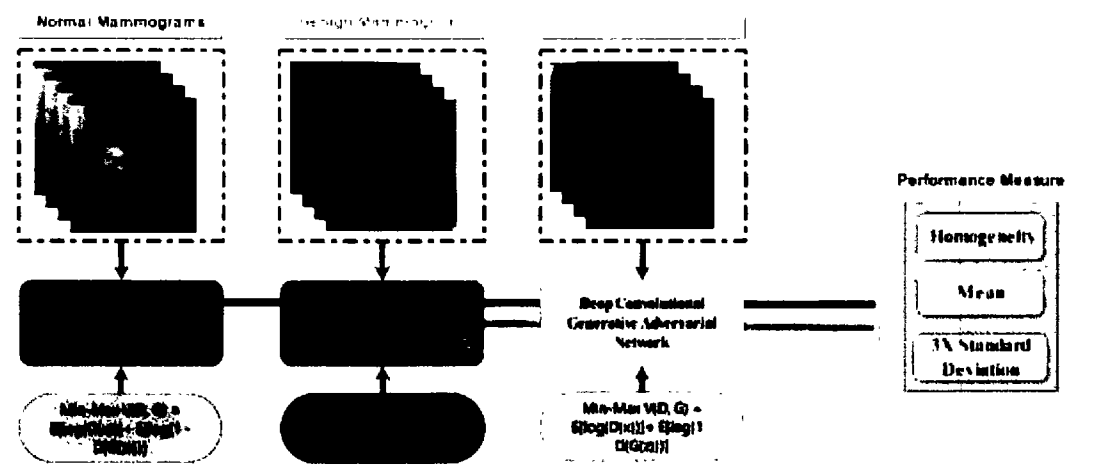


Figure 3.2: Methodology at Glance

3.2.1 Data Collection

The experiments in this study use a benchmark dataset called Digital Database for Screening Mammography (DDSM) [89]. A larger set of 2,620 DICOM-formatted digital mammograms from 262 patients make up the DDSM dataset. The dataset comprises 695 normal mammograms, 141 benign mammograms that did not require a callback, 870 benign mammograms, and 914 malignant mammograms. The ground truth data connected to every mammography picture containing questionable lesions is included in the description of DDSM. There are cases of both benign and malignant breast abnormalities within this spectrum.

3.2.2 Data Preparation

Preparing appropriate input data for the model is essential to ensure consistency and clinical relevance. Medical images, including mammograms, can be susceptible to noise and artifacts that might affect the quality of the training data. A denoising algorithm, such as median filtering or wavelet denoising, is applied to the mammogram images. This denoising process effectively reduces noise while preserving diagnostic features, resulting in cleaner images for training. The mammograms were resized to the same size and format to feed data to the network.

### 3.2.3 DCGAN Architecture

The architecture of the DCGAN [90] plays a pivotal role in generating realistic mammogram images. Its general architecture is shown in Figure 3.3. The process begins by sampling random latent vectors,  $z$ , from a normal distribution  $P_z(z \sim N(0,1))$  during each training iteration. These vectors are normalized to the range  $[-1, 1]$  and fed into the Generator ( $G$ ) to produce a batch of synthetic images ( $G(z)$ ). Both real and synthetic images are subsequently passed through the Discriminator ( $D$ ), which evaluates their authenticity by assigning a probability score. The loss for the Discriminator ( $LD$ ) is calculated based on its ability to correctly distinguish real images from synthetic ones, and the Discriminator's parameters are updated accordingly. Likewise, the Generator's loss ( $LG$ ) is determined by how well it can deceive the Discriminator into classifying the synthetic images as real. Throughout the training process, backpropagation is applied to iteratively update both the Generator and Discriminator, improving the Generator's ability to create realistic mammograms and enhancing the Discriminator's capacity to differentiate between real and synthetic images.

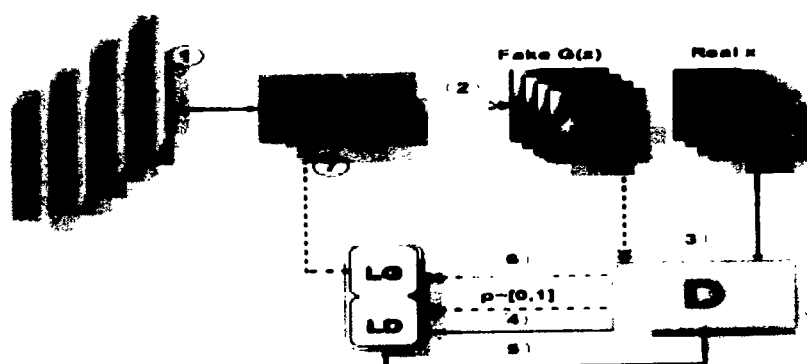


Figure 3.3: General Architecture of DCGAN

### 3.2.4 Generator Network

A random noise vector is fed into the generator network, slowly converting it into synthetic mammogram images. It starts with convolutional layers and then adds non-linearity with batch normalization and ReLU activation functions. Skip connections preserve key features during the down sampling process; they were inspired by U-Net architectures. The generator's architecture is summarized in Table 3.1.

Table 3.1: Generator Network

| Layer No. | Layer Type          | Output Shape | Activation |
|-----------|---------------------|--------------|------------|
| 1         | Dense (Input)       | 512×4×4      | -          |
| 2         | Reshape             | 256×4×4      | -          |
| 3         | Conv2DTranspose     | 512×8×8      | ReLU       |
| 4         | Batch Normalization | 256×8×8      | -          |
| 5         | Conv2DTranspose     | 128×16×16    | ReLU       |
| 6         | Batch Normalization | 64×16×16     | -          |
| 7         | Conv2DTranspose     | 32×32×32     | ReLU       |
| 8         | Batch Normalization | 32×32×32     | -          |
| 9         | Conv2DTranspose     | 3×64×64      | Tanh       |

3.2.5 Discriminator Network

The discriminator network's main job is identifying and distinguishing between real mammogram images and ones made artificially. Convolutional layers comprise architecture, followed by batch normalization and non-linear LeakyReLU activation functions. Table 3.2 provides an overview of the architecture of the discriminator.

Table 3.2: Discriminator Architecture

| Layer No. | Layer Type          | Output Shape | Activation |
|-----------|---------------------|--------------|------------|
| 1         | Conv2D (Input)      | 64×32×32     | LeakyReLU  |
| 2         | Conv2D              | 64×16×16     | LeakyReLU  |
| 3         | Batch Normalization | 64×16×16     | -          |
| 4         | Conv2D              | 128×8×8      | LeakyReLU  |
| 5         | Batch Normalization | 128×8×8      | -          |
| 6         | Conv2D              | 256×4×4      | LeakyReLU  |
| 7         | Batch Normalization | 256×4×4      | -          |
| 8         | Conv2D              | 512×4×4      | -          |
| 9         | Batch Normalization | 512×4×4      | LeakyReLU  |
| 10        | Flatten             | 1×1×1        | Sigmoid    |

The training simultaneously optimizes the generator and discriminator networks by using adversarial loss functions, such as binary cross-entropy or Wasserstein loss. The

Adam optimizer is used because of its robustness in handling non-stationary data and complex loss scenarios.

3.2.6 Training Process

The network's training process, which teaches the generator how to create realistic mammograms and the discriminator how to distinguish between real and fake images, is a critical component of this study. The following part describes the main elements of the training process, which include the tuning of the hyperparameters, the loss functions, and the convergence monitoring, as indicated in Table 3.3.

Table 3.3: Training Hyperparameters

| Hyperparameter                | Value  |
|-------------------------------|--------|
| Learning Rate (Generator)     | 0.0002 |
| Learning Rate (Discriminator) | 0.0002 |
| Batch Size                    | 64     |
| Number of Epochs              | 100    |
| Optimizer                     | Adam   |
| Beta 1 (Adam)                 | 0.5    |
| Beta 2 (Adam)                 | 0.999  |

Adversarial training involves the generator and discriminator networks playing a two-player minimax game. At the same time, the discriminator strives for greater accuracy in distinguishing between real and fake images. Meanwhile, the generator network aims to limit the discriminator network's ability to distinguish between real and synthetic mammogram images.

3.2.7 Loss Function

During training, DCGAN’s generator and discriminator primarily rely on binary cross-entropy loss. This loss quantifies the difference in ground truth and predicted labels for the discriminator's real/fake classification. During adversarial training, the Wasserstein loss enhances gradient flow and stabilizes training.

### 3.3 Performance Evaluation

Assessing the network's output performance and quality is critical for understanding its effectiveness and clinical applicability. The evaluation metrics are presented here to help determine the quality of synthetic images. A class-based similarity assessment was used to assess the quality of the DCGAN-generated images. Outliers in each class were identified and eliminated as needed. The outliers were eliminated above or below the stipulated threshold, as shown in Figure 3.4.

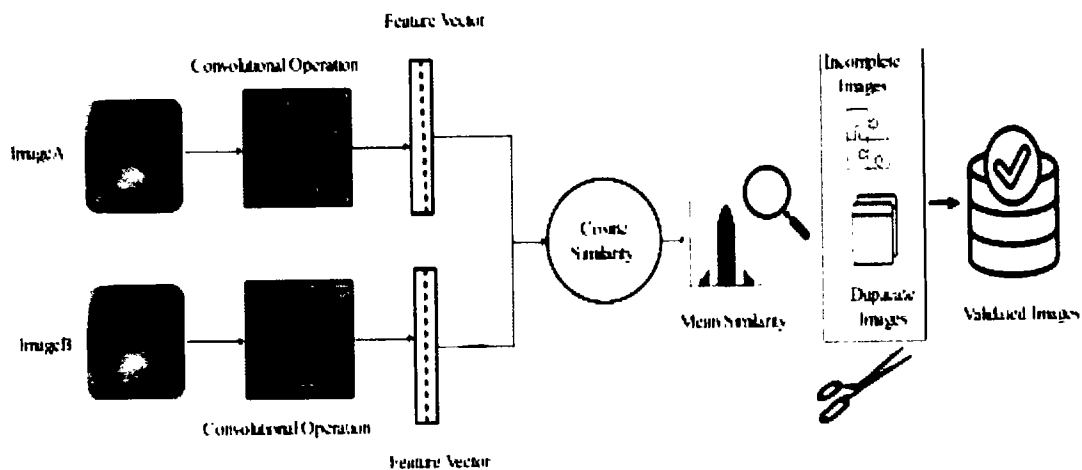


Figure 3.4: Methodology for Validation of Synthetic Images

The methodology for validating synthetic mammograms is shown in the following pseudo-code.

#### Pseudo-Code for Similarity Assessment and Outliers Removal

```

1: Procedure CosineSimilarity (A, B)
2:   return DotProduct(A, B) / (Norm(A) * Norm(B))
3: end Procedure

4: Procedure CalculateMeanVector (Vectors)
5:   return Average of Vectors along each dimension
6: end Procedure

7: Procedure CalculateStandardDeviation (Vectors)
8:   return StandardDeviation of Vectors along each dimension
9: end Procedure

10: Procedure CalculateCosineSimilarities (ImageVectors)
11:   CosineSimilarities ← Empty list
12:   for each Vector in ImageVectors do

```

| Pseudo-Code for Similarity Assessment and Outliers Removal |  |
|--|--|
| 13:  | Cosine $\leftarrow$ CosineSimilarity(Vector, CalculateMeanVector(ImageVectors))  |
| 14:  | Append Cosine to CosineSimilarities  |
| 15:  | end for  |
| 16:  | return CosineSimilarities  |
| 17:  | end Procedure  |
|  |  |
| 18:  | Procedure <b>CalculateThresholds</b> (CosineSimilarities)                        |
| 19:  | Mean $\leftarrow$ CalculateMean(CosineSimilarities)                              |
| 20:  | StdDev $\leftarrow$ CalculateStandardDeviation(CosineSimilarities)               |
| 21:  | ThresholdHigh $\leftarrow$ Mean + (3 * StdDev)                                   |
| 22:  | ThresholdLow $\leftarrow$ Mean - (3 * StdDev)                                    |
| 23:  | return ThresholdHigh, ThresholdLow   |
| 24:  | end procedure  |
|  |  |
| 25:  | Procedure <b>FilterImagesByCosineSimilarity</b> (ImageVectors)                   |
| 26:  | CosineSimilarities $\leftarrow$ CalculateCosineSimilarities(ImageVectors)        |
| 27:  | ThresholdHigh, ThresholdLow $\leftarrow$ CalculateThresholds(CosineSimilarities) |
| 28:  | FilteredImages $\leftarrow$ Empty list   |
| 29:  | for each Cosine in CosineSimilarities do   |
| 30:  | if Cosine $\geq$ ThresholdLow and Cosine $\leq$ ThresholdHigh then               |
| 31:  | Append corresponding ImageVector to FilteredImages                               |
| 32:  | end if   |
| 33:  | end for  |
| 34:  | return FilteredImages  |
| 35:  | end Procedure  |

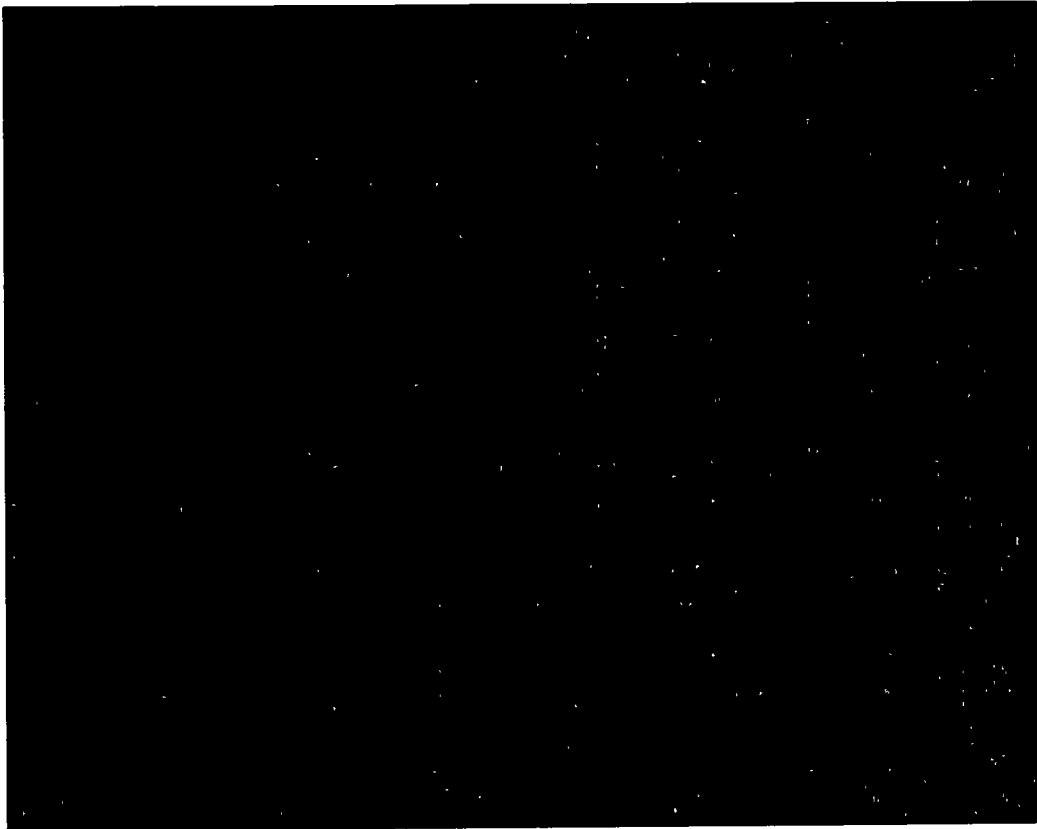
Combining these evaluation metrics ensures a robust and multidimensional assessment of the DCGAN-generated mammograms. By scrutinizing the synthetic images' structural, statistical, and diagnostic characteristics, the study gains valuable insights into the DCGAN's performance and its potential contribution to advancing medical imaging research and clinical practice.

3.4 Experiments and Results

Results of the study on the application of DCGAN for mammogram generation and the subsequent validation process are presented. The study focuses on three distinct classes of mammograms. It analyzes the mean homogeneity of each class and the distances of individual data points from their respective means using a statistical approach involving the three times standard deviation criterion.

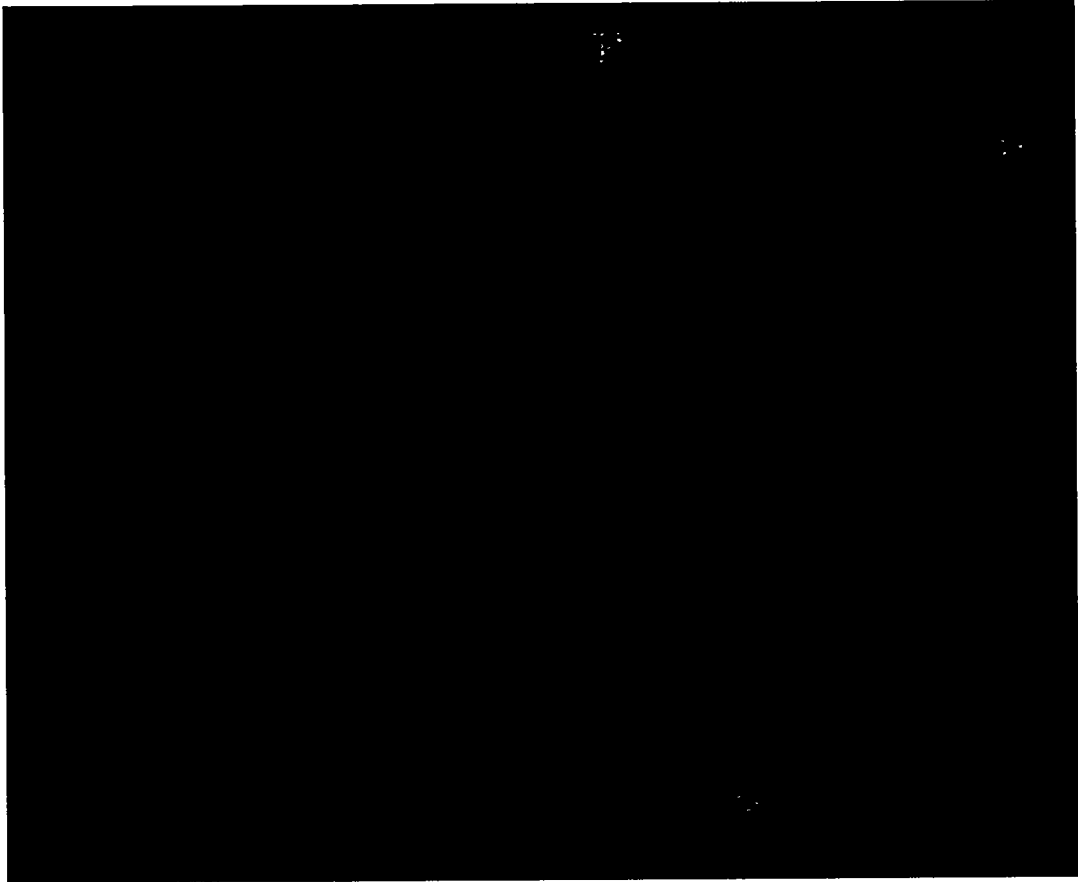
### 3.3.1 Synthetic Mammogram Generation

The network was first trained in images from three different classes of mammograms. The network was taught to make fake mammograms with many of the same features and characteristics as real ones. During training, the model was able to learn the patterns and structures that were unique to each class. This gave it the ability to make high-quality fake mammograms. The model generated 600 images for each class during the entire training over 100 epochs.



**Figure 3.5: Synthesized images from epoch 0 during training of DCGAN**

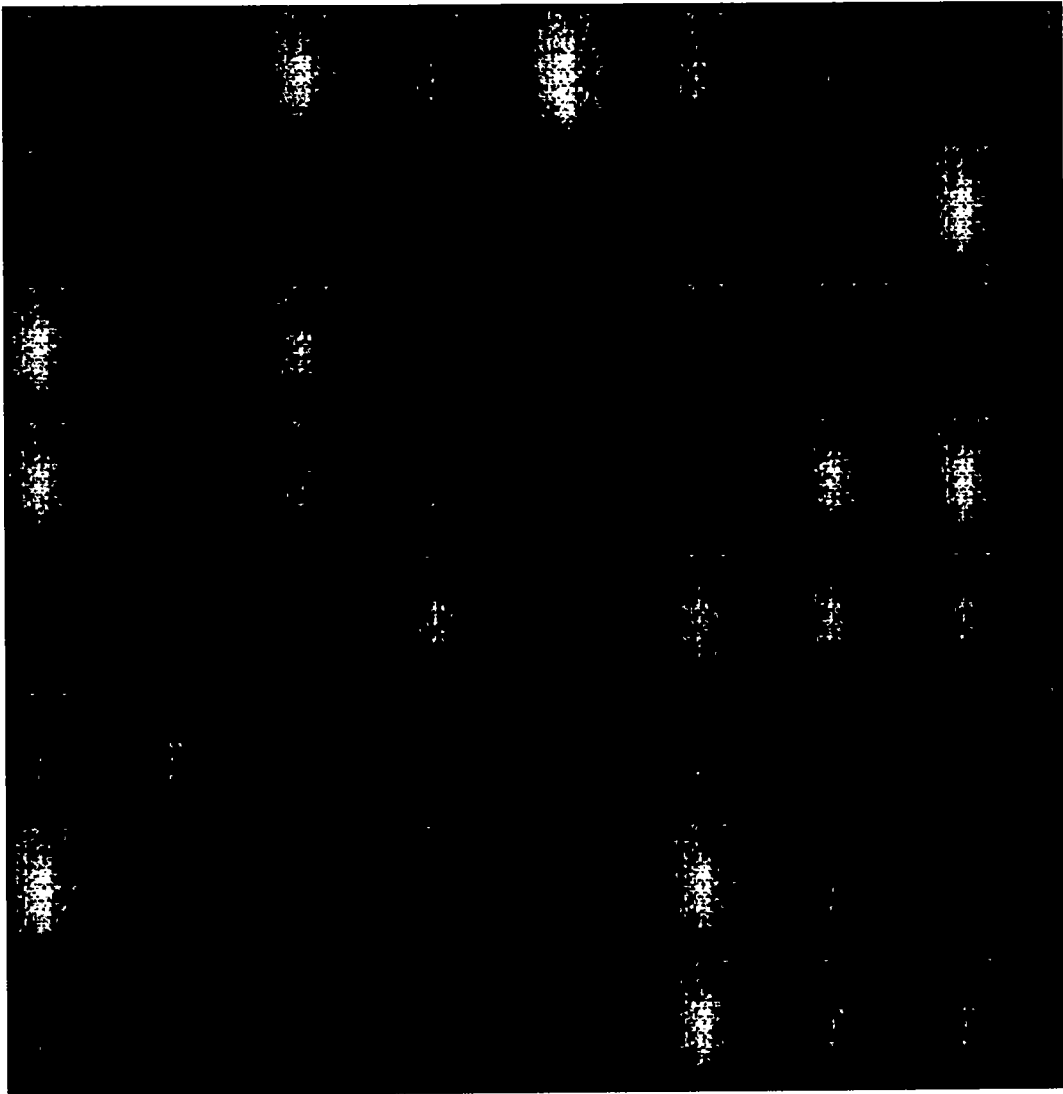
Figure 3.5 shows the initial images generated by the DCGAN from the input of random noise. At this stage, the DCGAN's generator starts with random initialization, producing patterns without any discernible structure indicative of the target data distribution. The generator attempts to map the data from noise, and as such, they do not yet resemble the desired output.



**Figure 3.6: Synthesized images from epoch 1 during training**

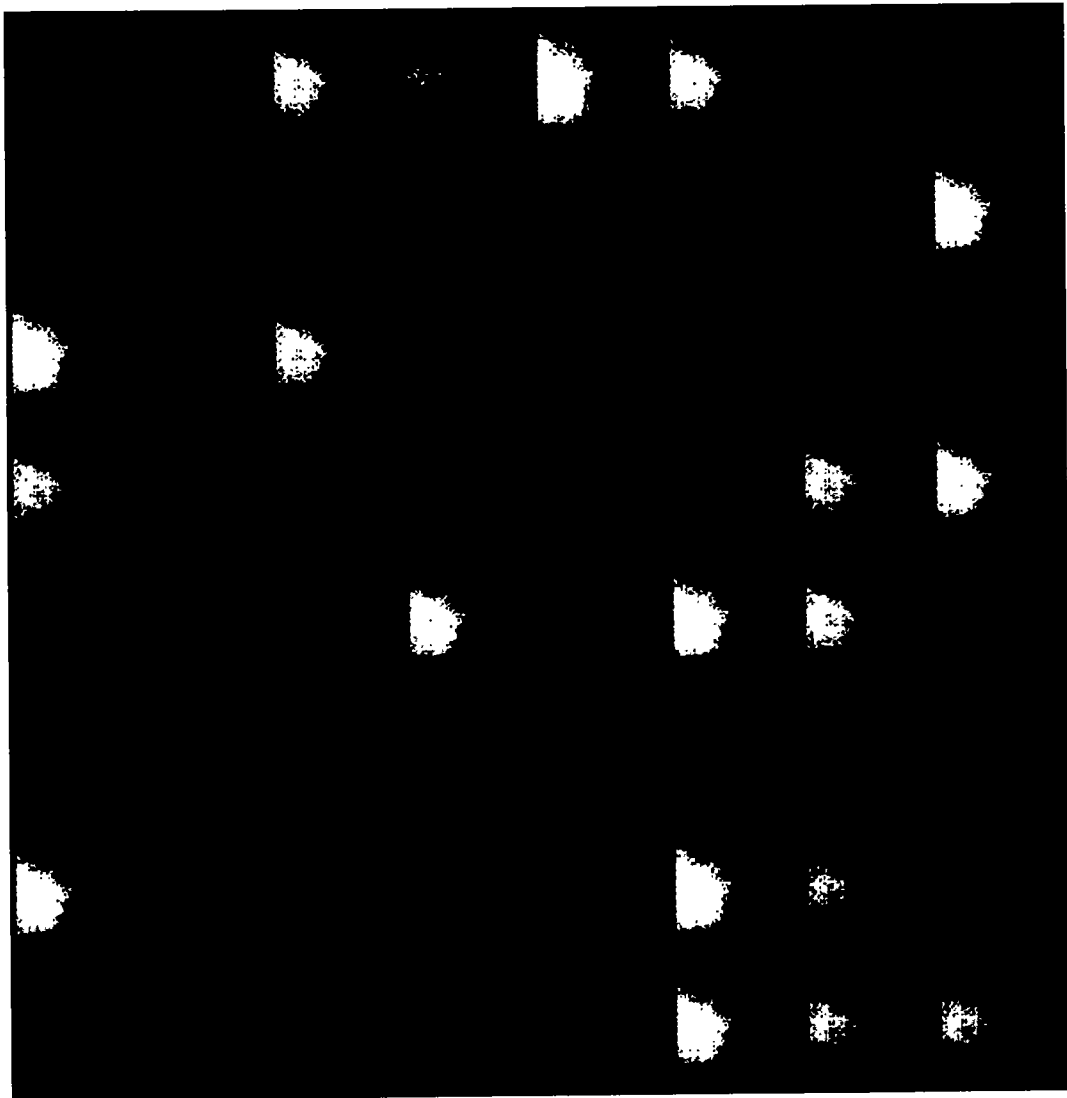
Figure 3.6 represents the images generated during the initial phases from epoch 1 during the training. The images show random noise at this stage. While still in the early training phase, the patterns exhibit slightly more structure than the initial epoch, suggesting the beginning of feature learning. Despite the emergent patterns, the images remain abstract and do not yet show identifiable objects or forms the network is expected to generate. This is typical of the DCGAN training progression, where the generator incrementally improves as it learns to create more complex features from the training dataset.





**Figure 3.7: Synthesized images from epoch 2**

Figure 3.7 shows synthesized mammograms generated from epoch 2 of the training process. The images are now slightly better than earlier images but still not recognizable. Notable progress from random noise is now observed, with the emergence of localized areas of coherence suggesting the network's initial understanding of patterns. These formations, while primitive and not yet resembling any defined shapes, indicate the network is beginning to learn the structural aspects of data distribution. This is a common stage in DCGAN training, reflecting early learning and the gradual refinement towards generating recognizable images.



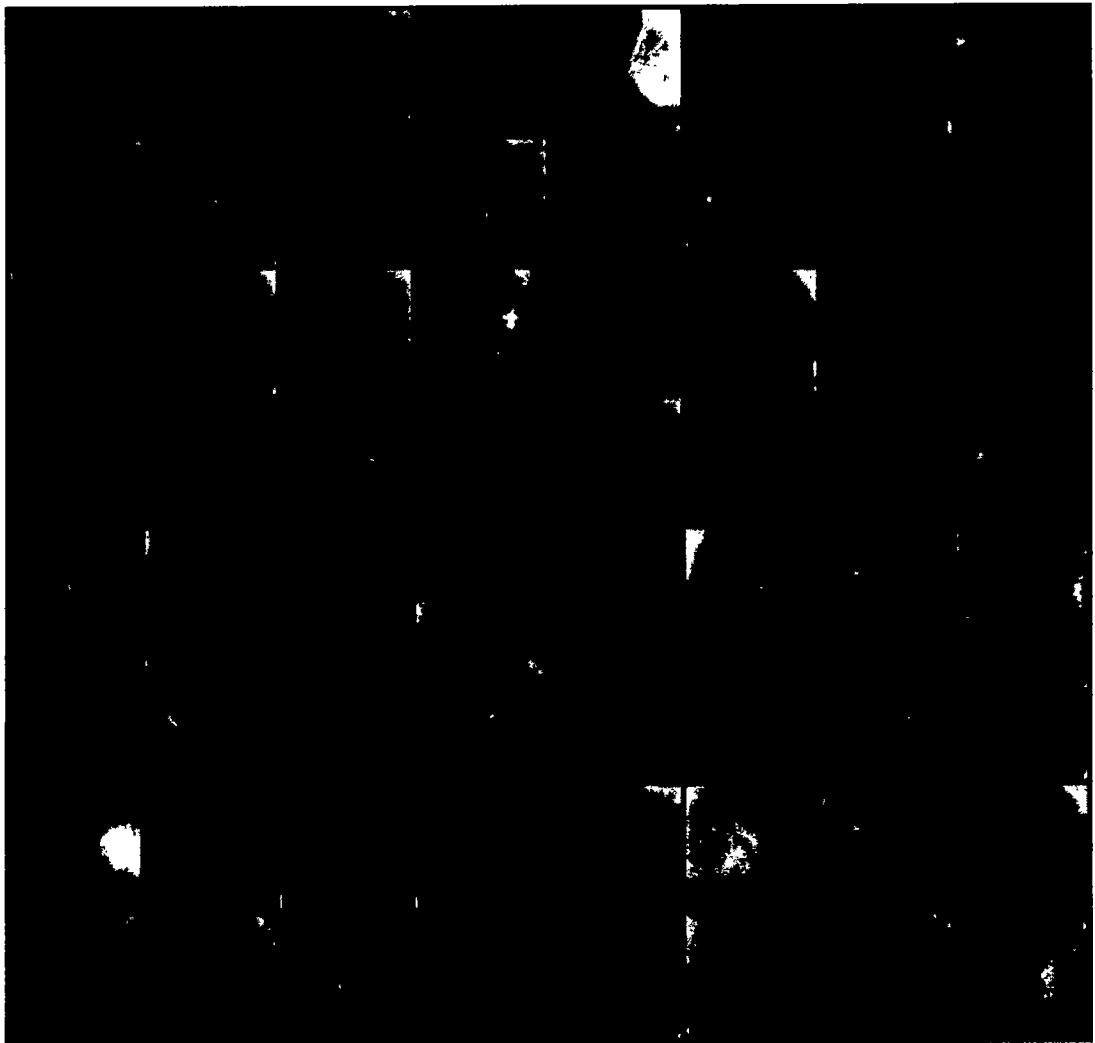
**Figure 3.8: Synthesized images from epoch 3 during training**

Figure 3.8 shows synthesized mammograms generated from epoch 3 of the training process. The images are shaping up. Now, there is a marked progression in the quality of the images, with more defined clusters of pixels beginning to take shape. Now, the generator network is starting to capture more complex features and patterns within the images from the training dataset, moving beyond the initial random noise and forming the rudiments of structure within the generated images.



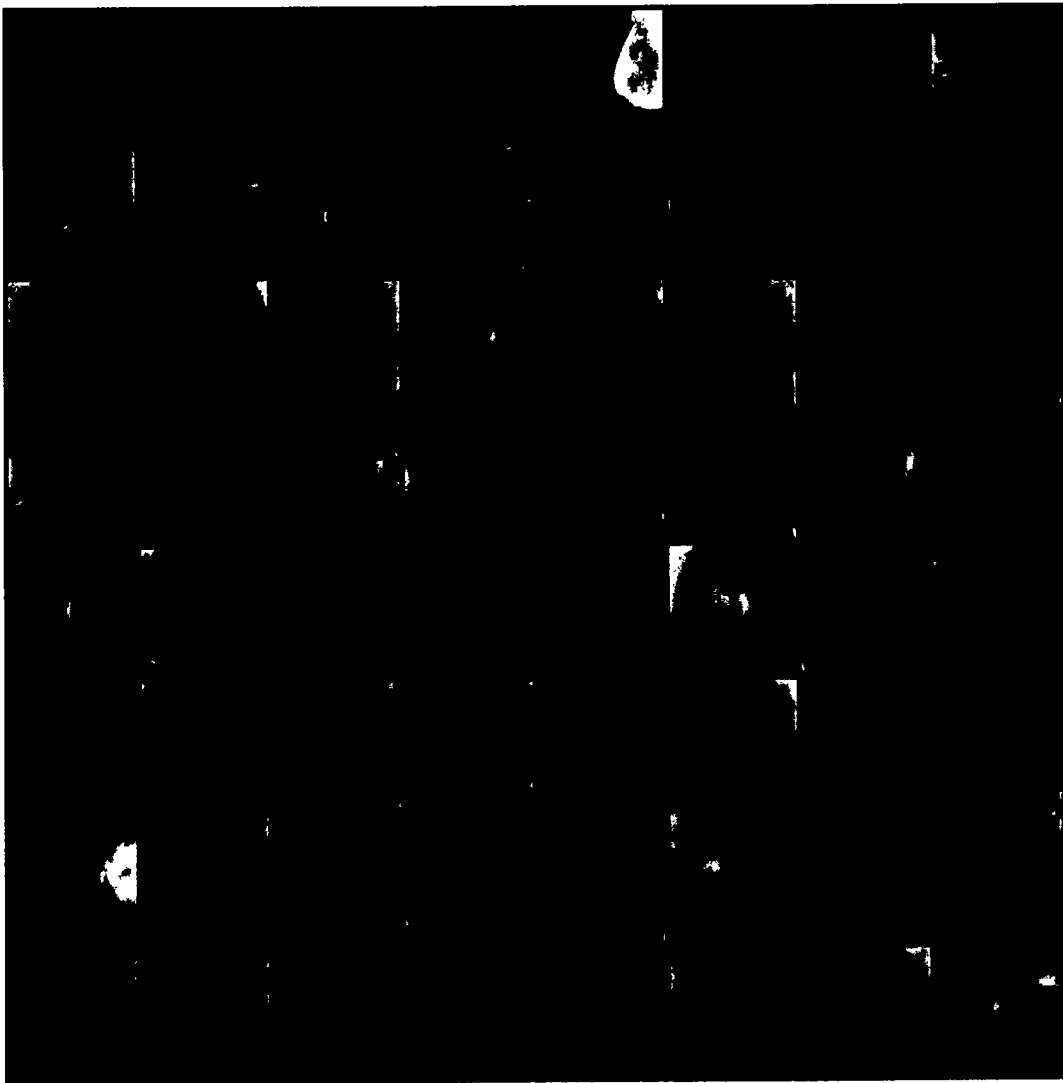
**Figure 3.9: Synthesized images from epoch 45 during training**

Figure 3.9 shows synthesized mammograms generated from epoch 45 of the training process of the DCGAN. The majority of the images now look like real mammograms. Now, the image shows a significant advancement in the learning process. The images now exhibit distinguishable features, reflecting the network's improved ability to synthesize representations closer to the actual data. This level of detail suggests that the DCGAN has effectively learned to capture and replicate the critical structures of the training dataset.



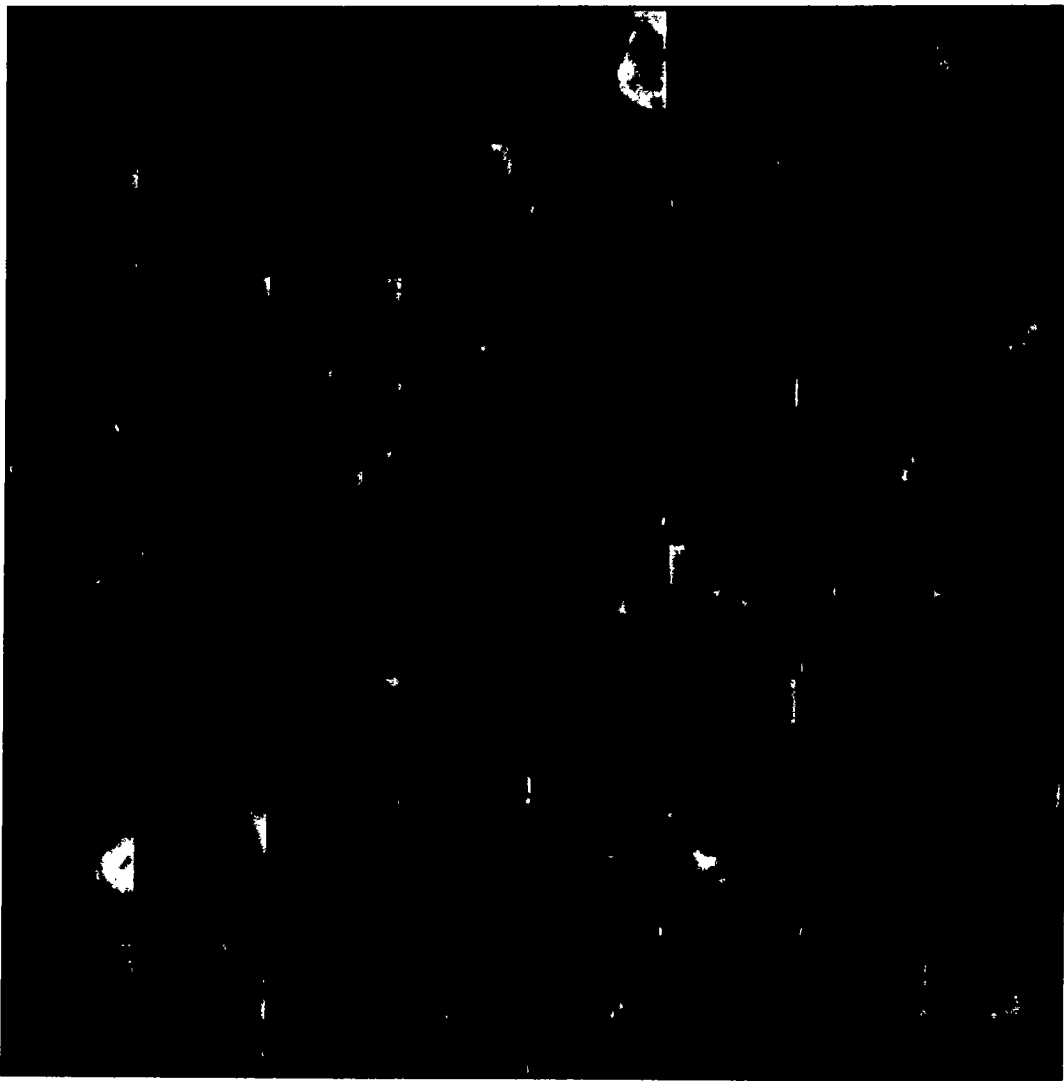
**Figure 3.10: Epoch 50 Images**

During the training of DCGAN, the synthesized images are now much better than all previous images during epoch 50, as shown in Figure 3.10. The images generated now show a clear progression in learning, with the visual output depicting discernible shapes and contrasts that suggest a closer approximation to actual data features. This significant improvement indicates DCGAN's advanced training stage, where the generator has refined its parameters to produce images with convincing details and contours.



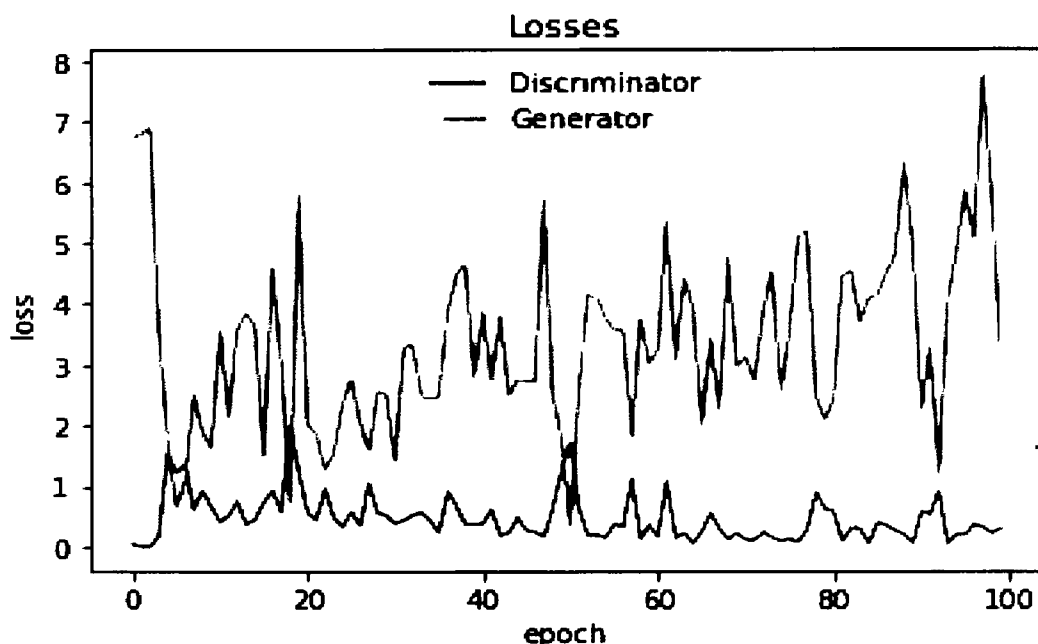
**Figure 3.11: Synthetic images during epoch 99**

During epoch 99, as shown in Figure 3.11, synthesized images are now in much better shape than ever before the previous epochs. The image produced here by DCGAN training indicates a mature stage of the generative model's learning. The images reveal distinct and clear features, reflecting the network's capability to synthesize complex structures that closely resemble the training dataset. This level of detail and definition in the images generated marks a significant step towards the network's convergence, demonstrating the efficacy of the DCGAN in capturing the nuances of the data it has been trained on.



**Figure 3.12: Images from epoch 100 during training**

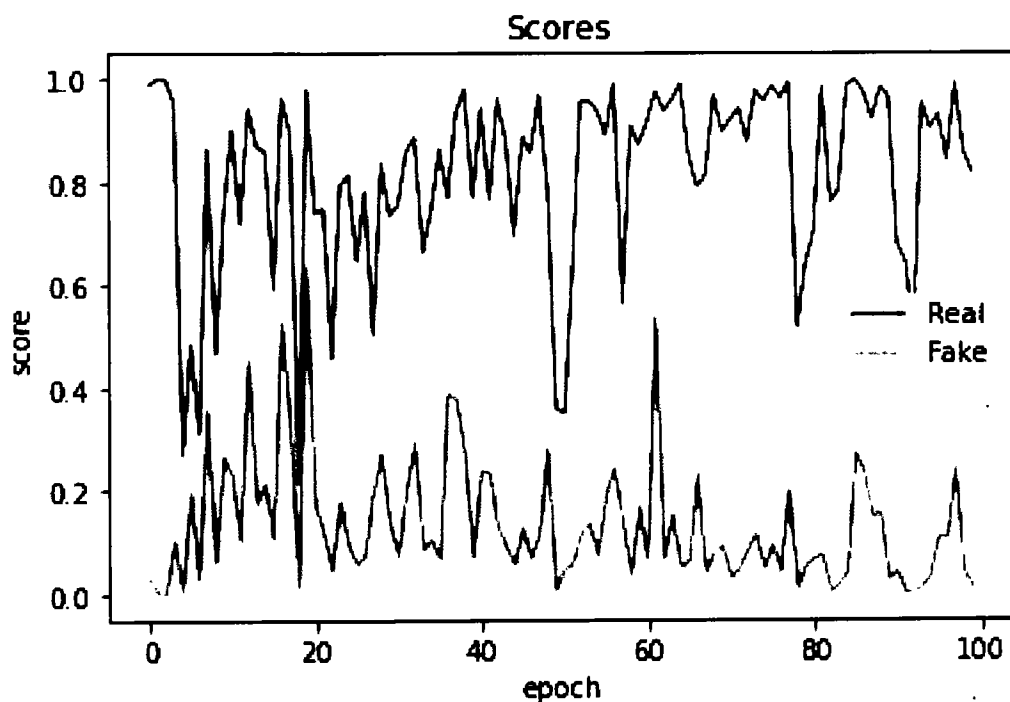
Epoch 100 was the last in the training process of the DCGAN, as presented in Figure 3.12. These are the finest images of the model during the entire training. The image from epoch 100 of DCGAN training demonstrates high sophistication in the generated images, where distinct and precise features are evident. This suggests that the generative model has effectively learned the training data distribution, achieving its goal of producing realistic and detailed synthetic images. Completing the DCGAN training cycle typically results in a model capable of generating images that can be difficult to distinguish from actual dataset samples, highlighting the success of the training process.



**Figure 3.13: Discriminator and Generator Network Losses During Training**

Figure 3.13 shows the losses of the Discriminator and Generator Network of the DCGAN model. The discriminator's goal is to classify real data from fake data correctly. As the discriminator network becomes more adept at differentiating between real and fake data, its loss decreases. The figure shows that the discriminator loss is low and stable, suggesting that the discriminator is performing its task with a consistent level of confidence. On the other hand, the generator's goal is to produce data that is indistinguishable from real data. Its loss decreases when the discriminator is more likely to mistake its generated data for real data. In the above figure, the generator's loss is volatile and tends to increase, indicating that the discriminator is getting better over time at distinguishing the fake data, or it could suggest that the generator is struggling to improve. The orange line shows significant volatility, spiking and dropping the loss throughout training. This is typical in the early stages of GAN training but can also indicate that the generator is having difficulty stabilizing. The general upward trend suggests that the discriminator is better at distinguishing real from fake over time, making the generator's task harder. Initially, the discriminator's loss decreases, suggesting it is rapidly learning to distinguish real mammograms from the synthetic ones created by the generator. The smooth line compared to the generator implies that the discriminator is learning with more stability. By the end of the 100 epochs, neither the generator nor the discriminator appears to have fully stabilized, which is common

in GANs. The smoother line for the discriminator loss indicates it may be approaching a point of stability faster than the generator.



**Figure 3.14: Real vs Fake Losses During Training**

Figure 3.14 shows the Real and Fake images of the model during training. The blue line representing the "Real" scores is quite jagged, with many spikes and dips. It suggests that the discriminator's confidence in recognizing real images is inconsistent. Ideally, we would want this line to be high, indicating strong and consistent recognition of real images, but the variability here implies the discriminator sometimes wavers in certainty. The orange line for the "Fake" scores depicts how often the fake images fool the discriminator. If the generator is doing well, this line should rise, reflecting that the fake images are becoming more convincing. If the line trends downward, the discriminator is better at telling them apart from real images.

Both lines fluctuate, which is common during network training as both networks are in a tug of war; as one improves, the other adapts. The graph does not show either line converging to a stable state, suggesting that the training process is still very dynamic and neither network has gotten the upper hand yet. This might indicate that the networks are still learning and require further training or parameter adjustments for better stability.



3.3.2 Mean Similarity

After generating synthetic mammograms, the mean similarity is calculated for each class, which provides insight into the consistency and similarity of the generated mammograms within that class.

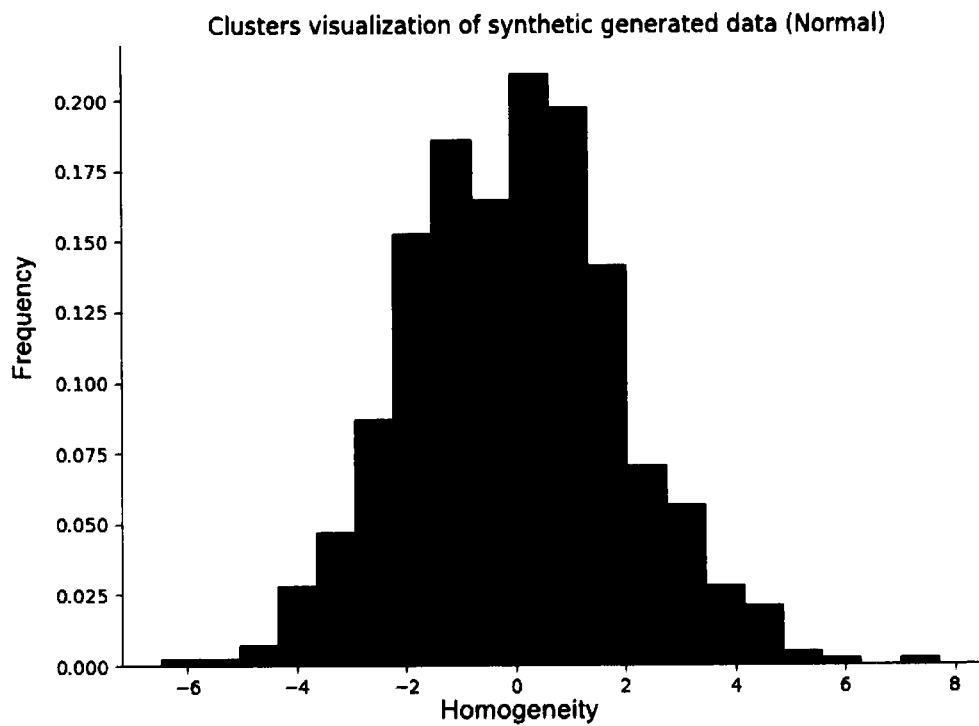


Figure 3.15: Cluster Visualization of Synthetic Normal Class Total Data

Figure 3.15 shows the total similarity and frequency of normal class mammograms. It shows that the data follows a normal distribution, with most values centered around zero. This indicates that similarity values most frequently occur near zero, with fewer occurrences of extremely low or high values.

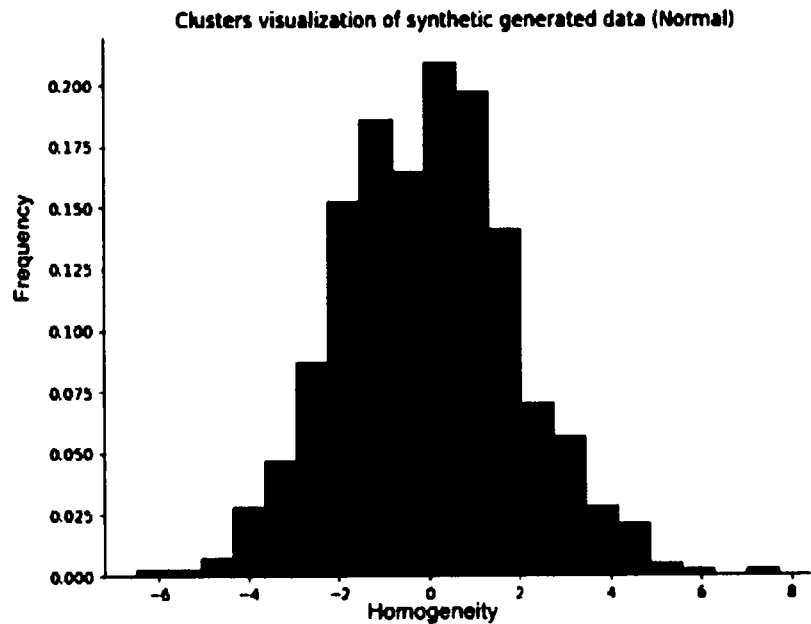


Figure 3.16: Cluster Visualization of Normal Class Data with Duplicate and Dissimilar Images

Figure 3.17 above shows the cluster visualization of synthetic normal class data without any outliers. Most data points cluster around the center, indicating that similar values near zero are the most common.

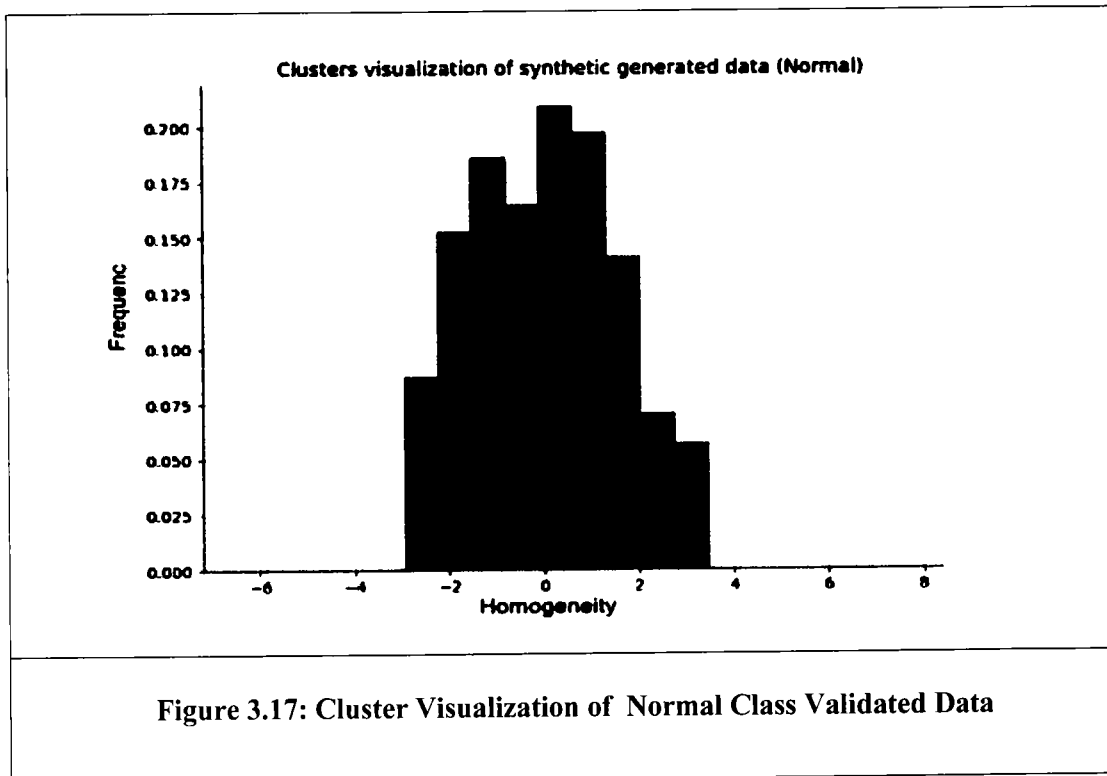


Figure 3.17 shows the cluster visualization of synthetic normal class data without the duplicate and dissimilar images above. The data is now without dissimilar and duplicate images.

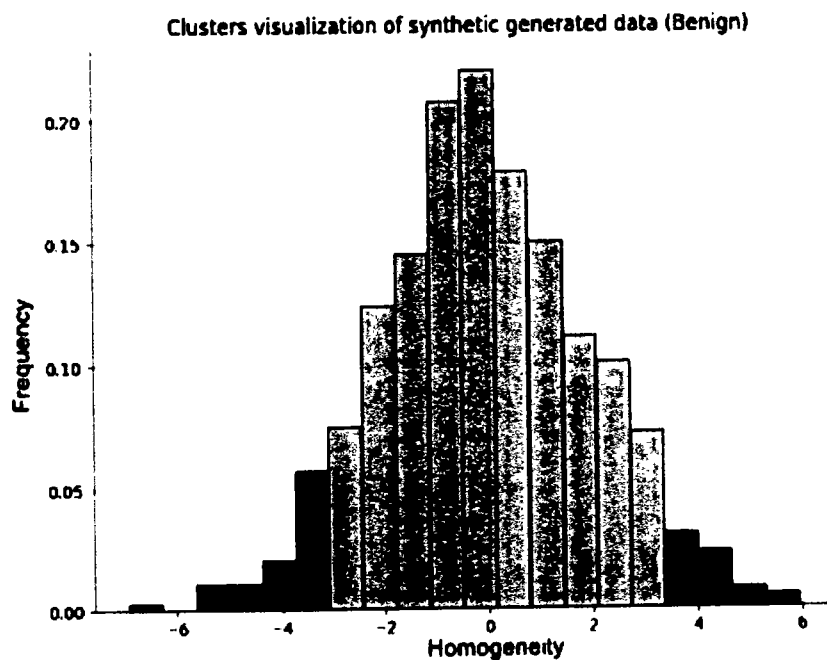


Figure 3.18 shows the cluster visualization of Benign class data with a few duplicate and dissimilar images on both sides of the scale, which are pointed with red color. Green bars represent the frequency of similar values.

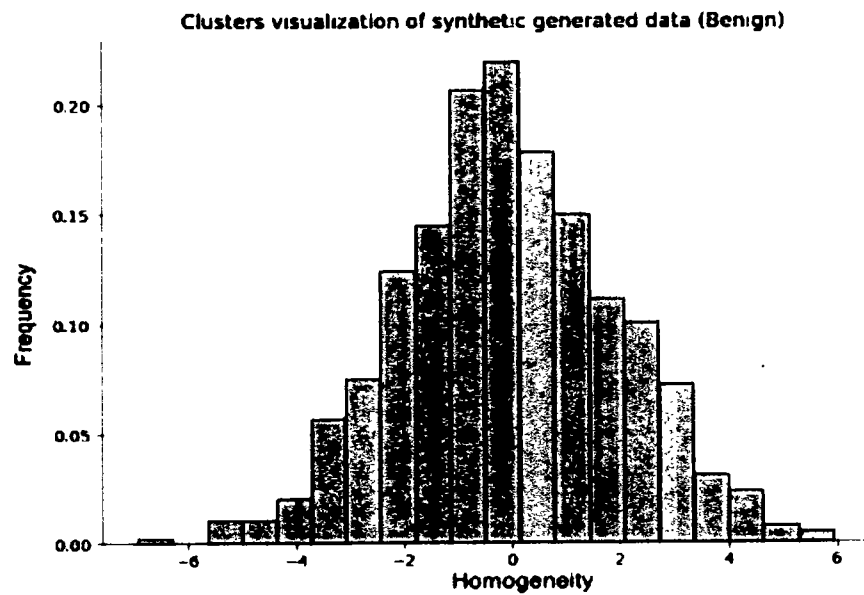


Figure 3.19: Cluster Visualization of Synthetic Benign Class All Data

Figure 3.19 shows the cluster visualization of synthetic Benign data generated by the DCGAN from both sides of the mean.

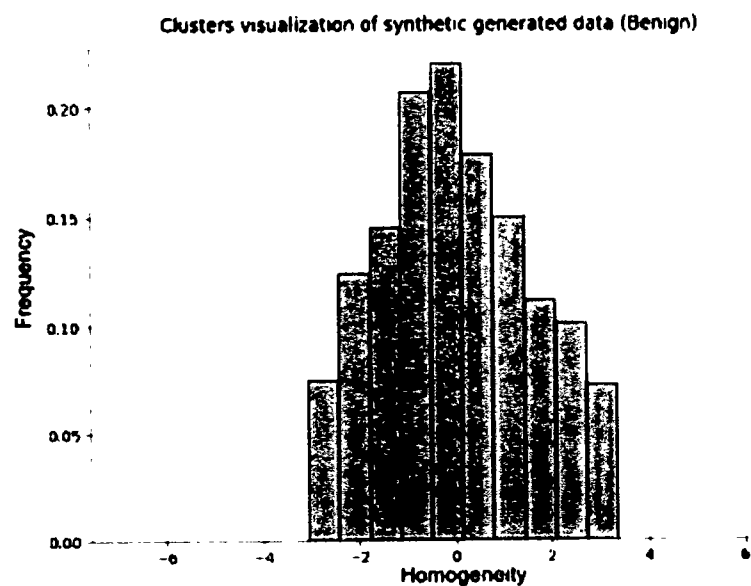


Figure 3.20: Cluster Visualization of Synthetic Benign Validated Class Data

Figure 3.20 shows the cluster visualization of synthetic Benign data generated by the DCGAN. The figure shows the images passed by the method for outlier removal. Duplicate and dissimilar images are removed from synthetic data.

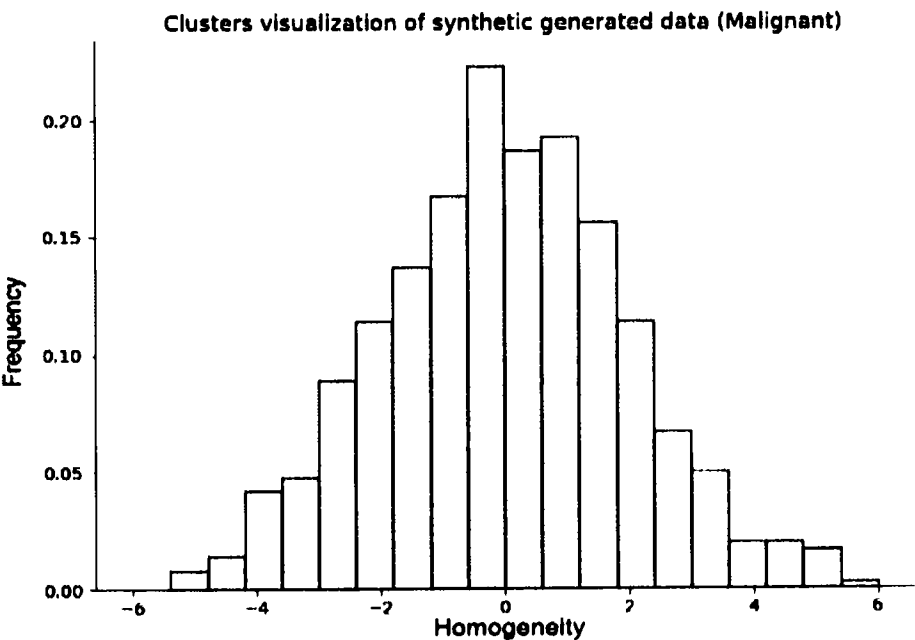


Figure 3.21: Cluster Visualization of Synthetic Malignant Class Data

Figure 3.21 shows the cluster visualization of synthetic Malignant data generated by the DCGAN.

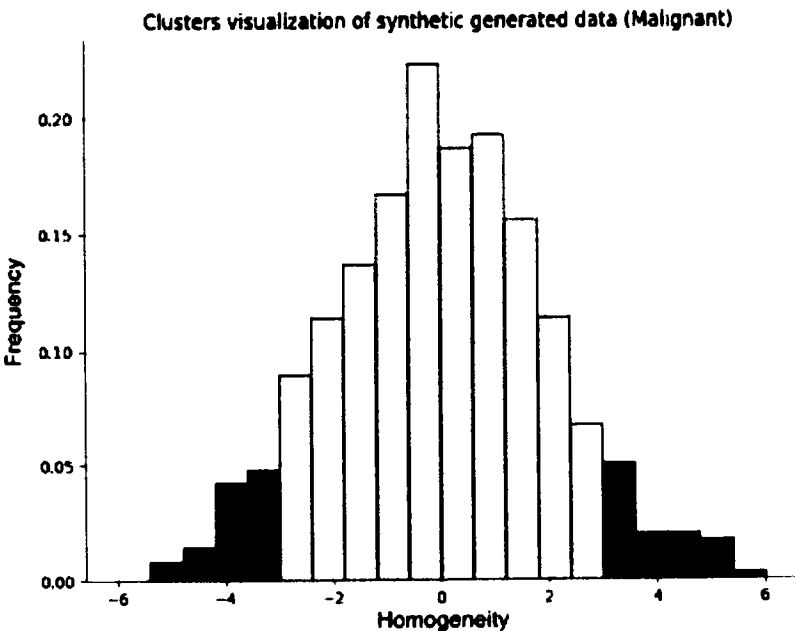
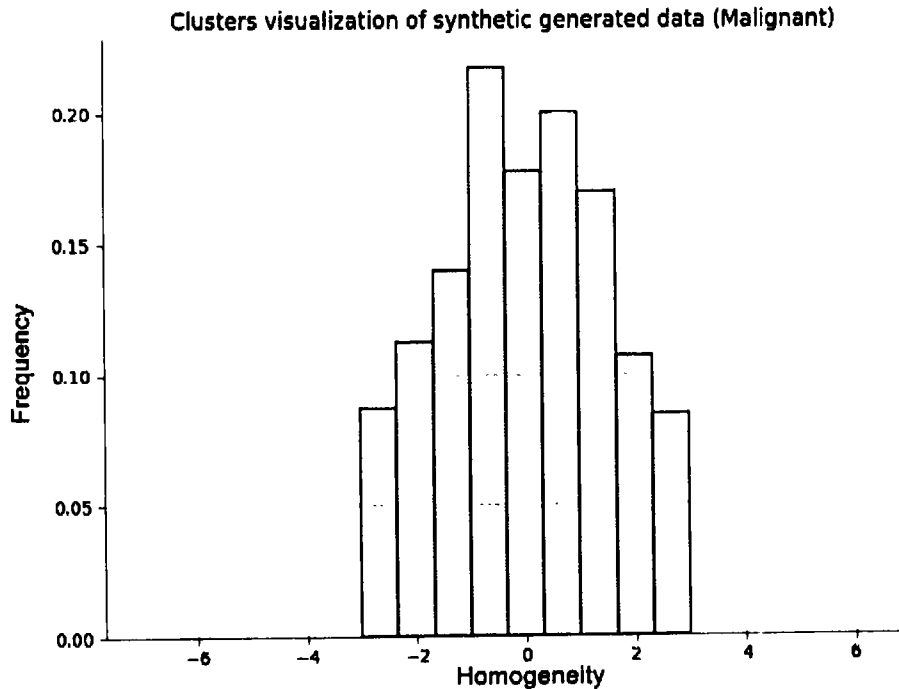


Figure 3.22 Visualization Malignant Class's Clusters with Duplicate and Dissimilar Images

In Figure 3.22 above, cluster visualization of synthetic Malignant data is shown with red duplicate and dissimilar images generated by the DCGAN. The duplicate and dissimilar images are present on the lower and higher ends of the similarity scale.



**Figure 3.23: Visualization of Malignant Validated Clusters**

The above Figure 3.23 shows the cluster of malignant data without duplicate and dissimilar mammogram images.

During validation of the synthetic images in a normal class, 95 images were declared either duplicate or dissimilar images out of a total of 600 images. In the benign class, 105 images were, while in the malignant class, 110 images were declared either duplicate or dissimilar images as per the similarity score.

### 3.3.3 Statistical Validation through Distance Calculation

The distance of each data point from its respective class mean is computed further to validate the quality and authenticity of the generated mammograms. This distance calculation involved utilizing the three times standard deviation criterion, which enabled us to quantify how much each generated mammogram deviated from the class mean. A more considerable distance value indicated a higher level of dissimilarity, whereas a smaller value suggested a closer resemblance to the mean.

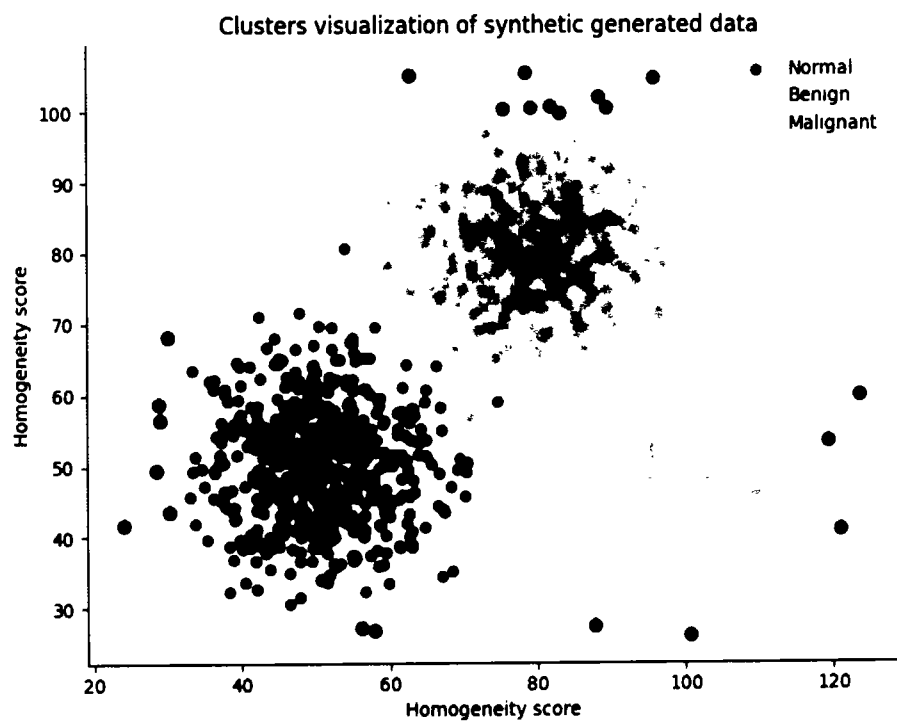
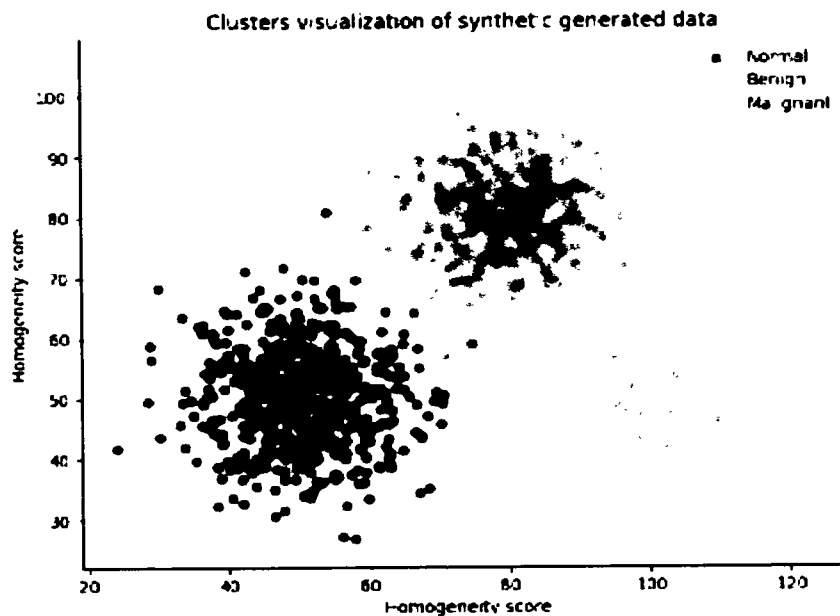


Figure 3.24: All Clusters Visualization of Synthetic Data with Duplicate and Dissimilar in Red

Figure 3.24 shows a scatter plot displaying clusters of synthetically generated data of three classes: Normal (blue), Benign (green), and Malignant (brown), with red dots indicating either duplicate or dissimilar images.



**Figure 3.25 All Clusters Visualization of Synthetic Data**

Figure 3.25 shows a plot for visualizing clusters within synthetically generated data. Three distinct groups are represented by different colors: blue for Normal, green for Benign, and brown for Malignant images. The dots' clustering suggests patterns within the data where similar values are grouped. The statistical validation approach is assessed to bolster the credibility of generated mammograms. This involved calculating the distance of each image from its class mean using the three times standard deviation criterion. More considerable distances indicated more significant dissimilarity, while smaller distances suggested better alignment with the class mean. In Figure 3.25, distinct and coherent clusters of similar data points were evident after removing duplicate or dissimilar images. This highlighted the effectiveness of our approach in forming meaningful clusters. Distance-based validation methods provided a robust means of quantifying the authenticity of the synthetic mammograms, improving clustering accuracy, and strengthening the reliability of the generated data for breast cancer imaging applications.



### 3.5 Validation of Synthetic Mammograms and Confidence Interval

#### Analysis

Confidence intervals (CI) provide a statistical measure that indicates the degree of confidence we have in the results obtained from our validation process. As stated earlier, 505 out of 600 in the normal class, 495 out of 600 in the benign class, and 490 out of 600 in the malignant class were validated.

So, for normal class validated mammograms  $p_{normal} = \frac{505}{600}$ ,  $p_{benign} = \frac{495}{600}$  and  $p_{malignant} = \frac{490}{600}$

The standard error (S.E) for a proportion is calculated using the formula is calculated.

$$S.E = \sqrt{\frac{p(1-p)}{n}} \quad (3.1)$$

Where  $p$  is the proportion valid synthetic mammograms and  $n$  is the total number of synthetic mammograms. Similarly, Confidence Interval (CI) is calculated using the formula given below:

$$CI = p \pm Z \times S.E \quad (3.2)$$

In the above equation  $Z$  is the  $Z$  – score corresponding the desired confidence level. For 95% confidence interval the value of  $Z = 1.96$

$p_{normal} = \frac{505}{600} \approx 0.8417$ ,  $p_{benign} = \frac{495}{600} \approx 0.825$  and  $p_{malignant} = \frac{490}{600} \approx 0.817$

Now Calculate S.E for all classes using equation (3.1).

$$S.E_{normal} = \sqrt{\frac{0.8417 \times (1-0.8417)}{600}} = \sqrt{\frac{0.8417 \times 0.1583}{600}} \approx 0.0149$$

$$S.E_{benign} = \sqrt{\frac{0.825 \times (1-0.825)}{600}} = \sqrt{\frac{0.825 \times 0.175}{600}} \approx 0.0155$$

$$S.E \text{ malignant} = \sqrt{\frac{0.817 \times (1-0.817)}{600}} = \sqrt{\frac{0.817 \times 0.183}{600}} \approx 0.0158$$

Now putting all values for each class in equation (3.2) as below:

For normal class confidence interval

$$CI \text{ normal} = 0.8417 \pm 1.96 \times 0.0149 = 0.8417 \pm 0.0292$$

$$CI \text{ normal} = [0.8125, 0.8709]$$

Similarly,

$$CI \text{ benign} = 0.825 \pm 1.96 \times 0.0155 = 0.8417 \pm 0.0304$$

$$CI \text{ benign} = [0.7946, 0.8554]$$

$$CI \text{ malignant} = 0.817 \pm 1.96 \times 0.0158 = 0.817 \pm 0.0310$$

$$CI \text{ malignant} = [0.7857, 0.8477]$$

The use of confidence intervals in the validation of synthetic mammograms provides a statistically sound approach to the assessment of the quality and accuracy of the data produced with the use of DCGANs. By estimating the confidence interval for the proportions of the valid synthetic mammograms in each class. Thus, allowing us to set up a reasonable measure of confidence in the synthetic data generated.

It can be observed that the confidence interval for normal mammogram is at 0.8125 to 0.8709. These figures are an indication that there is a high level of confidence that the true proportion of valid normal synthetic mammograms is within this range. Likewise, the benign class results in the confidence interval of 0.7946 to 0.8554 while for the malignant class, the confidence interval is 0.7857 to 0.8477. These intervals prove that the synthetic mammograms created are statistically similar to the real mammograms and therefore confirm the efficiency of the DCGAN model in creating realistic and usable synthetic data. This validation process is important since it gives a quantitative confirmation that the synthetic data may be used in subsequent studies, especially in the creation and evaluation of deep learning models for breast cancer diagnosis.

### 3.6 Validation from Human Experts

Considering how realistic some of the DCGAN-generated images look, we asked five radiologists with sufficient experience in radiology and mammography to classify synthetic and real images. Each radiologist was shown 240 images of a 50/50 mixture of real and synthetic images and was asked to rank them based only on their visual appreciation. The experts achieved a mean confidence score of 67%. In this review phase, a group of radiologists with experience in mammography evaluated the synthetic and real mammogram images. This panel was carefully selected based on their clinical expertise and familiarity with mammographic interpretation. Their assessment aimed to determine the realism and diagnostic accuracy of the synthetic images, comparing them to actual mammograms. The comments from the radiologists gave important information about the clinical viability of synthetic images. They ensured they complied with the requirements for efficient diagnostic use in a clinical setting.

### 3.7 Chapter Summary

This chapter discusses the generation of synthetic mammograms through DCGAN and its subsequent validation. The primary aim is to mitigate the issue of data scarcity in breast cancer imaging by generating high-quality synthetic mammograms. The chapter discusses the importance of data augmentation in general and synthetic data in particular. The architecture of DCGAN and its tuning is discussed next in detail. Experiments show the effectiveness of the DCGAN-generated mammograms, with visual comparisons to real images highlighting their realism. Our second objective was to validate these synthetic mammograms. Synthetic images were statically validated using mean homogeneity and 3 x standard deviation. Mammograms beyond a specific threshold were either duplicate images or dissimilar and were subsequently discarded from the dataset. The chapter emphasizes the significant contribution of using DCGANs for synthetic data generation, which can lead to improved training datasets and more dependable DL models in medical imaging.

## Chapter 4

### Dual-View Mammogram Classification

Mammography image classification is essential to the early diagnosis and detection of breast cancer. Images from mammograms are classified using a variety of models. The research community's goal in image-based diagnosis is to close the knowledge gap between clinical application and medical imaging technology by providing a thorough overview of the process of classifying mammograms, the difficulties encountered, and the implications for treatment outcomes and patient care. This chapter goes into great detail about the mammography classification process.

#### 4.1 Introduction to Dual Pathway Classification

In 2020, there were over 2.3 million cases of BC, the most common type of cancer worldwide, and over 685,000 deaths from BC [91]. It remains one of the most common forms of this illness [92], especially in women around the world, and has a significant impact on the total number of deaths. BC is characterized by abnormal cell division, resulting in breast tissue tumors. It is typically felt as a painless lump or a thickened area of the breast [93]. Mammography analysis is crucial for the early detection and diagnosis of breast cancer, one of the most prevalent and potentially deadly cancers that affect women globally. A low-dose X-ray image of the breast tissue is called a mammogram. It allows healthcare professionals to identify abnormalities, such as tumors or calcifications, often before they can be felt or cause noticeable symptoms [94]. Early detection through mammography has been proven to significantly increase the chances of successful treatment and improved patient outcomes.

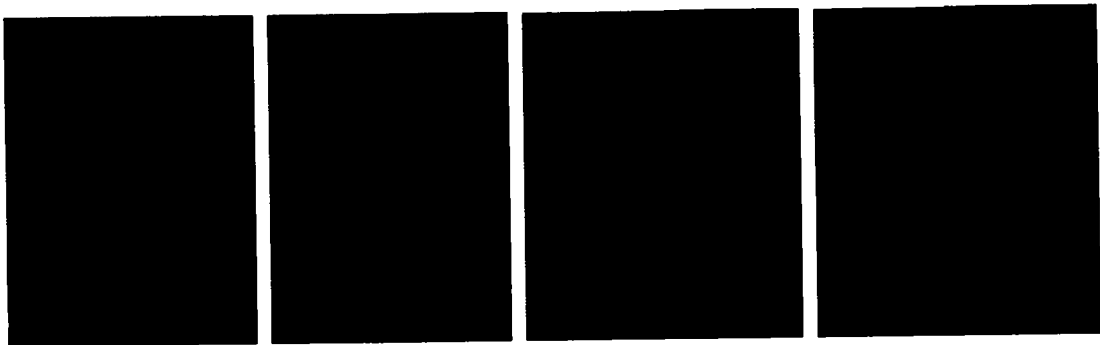
However, the interpretation of mammograms poses several challenges. One critical challenge is the presence of false positives and false negatives in the classification process [95]. When a mammogram is classified as positive (indicating potential cancer) even though no cancerous growth is present. False positives can lead to unnecessary anxiety, additional tests, and potentially invasive procedures, causing psychological stress and increased healthcare costs for patients [96]. On the other hand, false negatives occur when a mammogram is classified as negative, failing to identify an existing

cancerous lesion. False negatives can delay timely intervention, allowing the disease to progress and potentially leading to more aggressive treatments and reduced survival rates [97].

Because it directly affects patient well-being, the use of healthcare resources, and the overall success of breast cancer screening programs, it is crucial to lower false positives and false negatives in mammogram analysis. Innovative methods that improve the accuracy and reliability of mammogram classification are needed to address these issues [94]. The task of mammogram classification presents a formidable challenge. Utilizing a single classifier gives certain constraints, such as the potential for bias and overfitting, impeding the ability to effectively generalize novel and unseen data [84]. It is possible that the alone classifier may not possess the capacity to fully comprehend the intricate nuances included within the mammographic data [98]. A standalone classifier may face imbalanced data challenges in certain instances, resulting in skewed predictions [99]. Furthermore, the performance of a singular classifier is significantly influenced by its architectural design and initial configuration. If the algorithm becomes trapped in a local minimum or is inadequately initialized, it may yield unsatisfactory outcomes [100]. Likewise, a single classifier may exhibit sensitivity towards noise or outliers in the dataset, resulting in erroneous categorizations [101]. Breast cancer is a highly prevalent neoplasm affecting women globally, and the utilization of mammography as a means of early detection continues to be of paramount importance in mitigating mortality rates. The process of interpreting mammography poses considerable difficulties owing to the intricate nature of breast tissue and the nuanced characteristics of malignant tumors. Computer-aided diagnostic systems have exhibited the potential to augment the precision of radiologists, whereas DL models have indicated notable achievements in diverse medical image-processing tasks.

The main objective of this study is to create and implement EfficientNet architectures with multiple pathways specifically optimized for classifying mammograms. This approach utilizes architecture's ability to efficiently capture local and global information. This study aims to assess the effectiveness of multi-pathway models in dealing with challenges related to inadequate data, uneven distribution of classes, and variations in mammography presentations.

The need for better early cancer detection is what drives this research. The complex interpretation of mammography, a crucial tool in early diagnosis, creates difficulties. The study presents a novel strategy motivated by DL and diagnostics developments. It uses two perspectives and strikes a balance between recall and precision. The primary motivation for this research is the necessity to improve the accuracy of breast cancer diagnosis. This is achieved by integrating advanced DL methodologies with the extensive collection of benchmarks' mammographic RSNA dataset, as shown in Figure 4.1.



**Figure 4.1: Sample Mammograms from RSNA Benchmark Dataset**

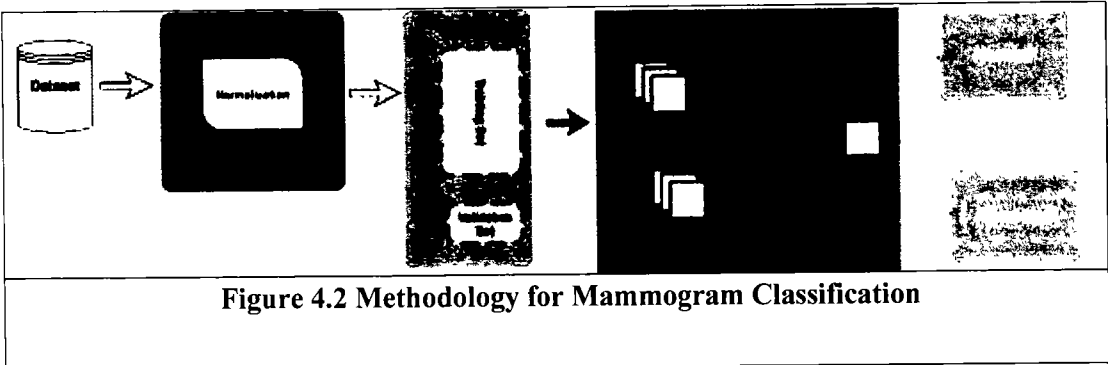
This study explores the capabilities of multi-pathway EfficientNet architectures, which possess the distinct capacity to capture detailed features and adjust to different scales of visual input. This research endeavor has the potential to address the constraints associated with conventional single models, therefore facilitating the development of more precise and dependable classification. Below are the contributions of this study.

- The EfficientViewNet architecture effectively mitigates modality shift by leveraging two pathways. The distinctive architecture consistently enhances the performance of the model.
- The model effectively optimizes its learning rate for the corresponding perspectives.
- The architecture achieved superior results and reduced computational complexity compared to other variants of deep neural networks.

## **4.2 Methodology for Dual Pathway EfficientViewNet Model**

This section describes the EfficientViewNet-based mammography classification process. Mammographic images of the same breast are typically captured from two

standard angles: the CC view from above and the MLO view from the side, as illustrated in Figure 4.2. The essential steps are to prepare the data, create the model's architecture, and fine-tune it. The methodology's key components, learning rate analysis and comparative evaluation with state-of-the-art techniques enable a thorough evaluation of its effectiveness.



4.2.1 Dataset Description and Preprocessing

In this section, a brief explanation of the dataset is presented here. The study is based on a portion of the mammography dataset from the Radiological Society of North America (RSNA). [102], a widely recognized collection of mammograms encompassing diverse cases from various clinical settings. 31 radiologic subspecialties are represented by the nonprofit organization RSNA, which also represents 145 countries. Typically, each patient is provided with four images, consisting of two lateral left images and two lateral right images per view. This benchmark dataset includes various complexity levels and instances with different breast densities, lesion types, and size distributions. The robustness of the model depends on this diversity for it to handle the real-world situations that arise in clinical practice. A series of preprocessing steps were executed on the dataset to ensure consistent and effective training.

EfficientNet architectures inherently require images of specific dimensions. Thus, all mammogram images were uniformly resized to match the input dimensions of the respective EfficientNet pathways dedicated to CC and MLO views. This step ensures compatibility between the model architecture and the data. Normalization is a crucial step in bringing pixel values to a consistent scale. The study facilitates smoother convergence during training by entering the pixel values around zero mean and scaling them to a standard deviation of one. It mitigates the impact of variations in pixel

intensity. Validation and training sets were created from the dataset. The validation set helps with hyperparameter tuning. Techniques such as oversampling, under sampling, or a combination of both were employed to tackle the class imbalance issue in this dataset to ensure that the model does not disproportionately favor the majority class during training.

#### 4.2.2 Architecture of the EfficientViewNet

This section highlights the architecture of the EfficientViewNet model with its suitability for the mammogram classification task. EfficientViewNet comprises two subnetworks, CC-EfficientNet and MLO-EfficientNet, presented in Figure 4.3.

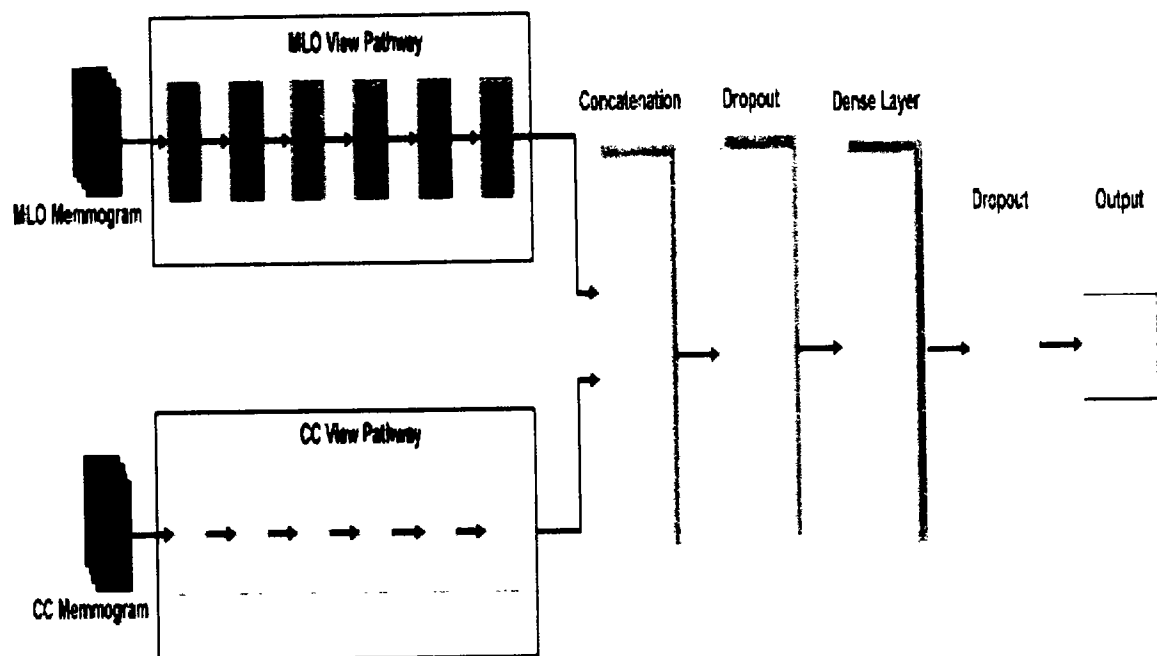


Figure 4.3: Architecture of the EfficientViewNet

#### 4.2.3 CC View Pathway

The CC view pathway is designed to extract relevant features from CC mammogram images through a dedicated branch of the dual pathway EfficientNet-based model. The architecture of the CC view pathway is tailored to the unique characteristics of CC view mammograms, optimizing its ability to detect subtle patterns indicative of breast abnormalities. The network architecture of the CC view pathway is based on an



EfficientNet model, trained on a large dataset to leverage its feature extraction capabilities. The architecture is represented in equation (4.1).

$$CC\ Pathway(X) = Conv_{F_1} * BN * ReLU * ... * Conv_{F_N} * BN * ReLU * Drop \quad (4.1)$$

Where  $N$  is given in the above equations, it is the total number of layers in the  $CC$  view Pathway.  $Conv$  represents convolutional operation.  $ReLU$  represents the Rectified Linear Unit activation function.  $BN$  represents Batch Normalization.  $F$  represents the number of filters in the convolutional layer, whereas  $D$  represents the dropout rate.

The configuration parameters of the  $CC$  view pathway, such as learning rate, optimizer, and batch sizes, are tuned through preliminary experiments and validation. An overview of the significant configuration parameters used during training is provided in Table 4.1.

Table 4.1: CC View Pathway Configuration Parameters

| Configuration Parameter | Value         |
|-------------------------|---------------|
| Learning Rate           | 0.001         |
| Batch Size              | 32            |
| Optimizer               | Adam          |
| Loss Function           | Cross-Entropy |
| Dropout Rate            | 0.3           |

A multi-step process is used to train the  $CC$  view pathway to improve the model's ability to classify  $CC$  view mammograms accurately. The steps are described as follows:

The training pipeline receives preprocessed  $CC$  view mammogram images and their corresponding labels. Using the loaded data, the model is trained. A mathematical representation is given below.

$$D_{CC} = \{(X_{CC}, y_{CC})\}N_{cc} \quad (4.2)$$

Where  $X_{CC}$  represents preprocessed  $CC$  view mammogram images and  $y_{CC}$  their corresponding labels,  $N$  is the total number of samples in the dataset. To minimize the loss function, which in this context is Cross-Entropy Loss:

$$\mathcal{L}(W_{CC}; D_{CC}) = -\frac{1}{N} \sum_{i=1}^{N_{CC}} [y_{CC,i} \log(p(y_{CC,i} | x_{CC,i}; W_{CC})) + (1 - y_{CC,i}) \log(1 - p(y_{CC,i} | x_{CC,i}; W_{CC}))] \quad (4.3)$$

Where  $p(y_{CC,i} | x_{CC,i}; W_{CC})$  is the predicted probability of the true label  $y_{CC}$  given input  $X_{CC}$ .

$$W_{t+1} = W_{t-nt} - \nabla w_t \mathcal{L}(w_t; D) \quad (4.4)$$

Where  $\nabla w_t \mathcal{L}(w_t; D)$  represents the gradient of the loss with respect to the weights at time  $t$ .

Using the Adam optimizer to optimize the selected loss function (cross-entropy), the model iteratively modifies its weights during training. By causing some neurons to be randomly deactivated during training, dropout layers help to prevent overfitting. The model's performance is assessed using a distinct validation set not utilized during training. This enables us to evaluate the model's ability to generalize and make well-informed choices regarding adjustments to hyperparameters.

Based on validation performance, hyperparameters and model weights are fine-tuned to enhance classification accuracy. Learning rate schedules and early stopping mechanisms are applied to prevent overfitting and ensure optimal convergence. The final system classification  $\hat{y}$  can be modeled as.

$$\hat{y} = F_{final}(F_{CC}(x_{cc}; W_{cc}), F_{MLO}(X_{MLO}; W_{MLO}); W_{final}) \quad (4.5)$$

$F_{CC}$  and  $F_{MLO}$  represent the CC and MLO pathways.  $x_{cc}$  and  $X_{MLO}$  are the input mammograms from each pathway.  $W_{cc}$ ,  $W_{MLO}$  and  $W_{final}$  are the weights for the CC pathway, MLO pathway and final classification system.

The training process is iterated until the model performs satisfactorily on the validation set. The resulting CC view pathway model is then integrated with the MLO view pathway to create a mammogram classification system.

4.2.4 MLO View Pathway

The MLO view pathway is a pivotal component of this dual pathway EfficientNet-based classification model, designed to extract pertinent features from MLO mammogram images. This pathway captures distinct patterns and structures unique to MLO view mammograms, enabling the pathway for accurate breast cancer classification. The MLO view pathway is designed as its CC view counterpart.

$$MLOPathway(X) = Conv_{MLO\ F_{MLO}} * BN_{MLO} * ReLU_{MLO} * ... * Conv_{F_{MLO}} \quad (4.6)$$

$N_{MLO}$  is the total number of layers in MLO Pathway.  $Conv_{MLO}$  represents the convolutional operation.  $ReLU_{MLO}$  represents Rectified Linear Unit activation function.  $BN_{MLO}$  represents Batch Normalization where  $F_{MLO}$  represents the number of filters in the convolutional layer.

Key configuration parameters governing the MLO view pathway, such as the learning rate, batch size, and optimization algorithm, are determined through experimentation and validation. Table 4.2 outlines essential configuration parameters for the MLO view pathway.

Table 4.2: MLO View Pathway Configuration Parameters

| Configuration Parameter | Value         |
|-------------------------|---------------|
| Batch Size              | 32            |
| Optimizer               | Adam          |
| Learning Rate           | 0.001         |
| Dropout Rate            | 0.2           |
| Loss Function           | Cross-Entropy |

The training procedure of the MLO view pathway mirrors that of the CC pathway, albeit with adaptations specific to the unique characteristics of MLO mammogram images. The training process can be broken down into several steps:

Preprocessed MLO mammogram images and corresponding dataset labels are loaded for training. The model undergoes training by utilizing the Adam optimizer and minimizing the selected loss function, which is cross-entropy. Dropout layers are incorporated to mitigate overfitting concerns in the model’s training. Performance

monitoring is conducted using a separate validation set. This set provides insights into the model's generalization capabilities and guides hyperparameter adjustments.

$$D_{MLO} = \{(X_{MLO}, y_{MLO})\}_{i=1}^{N_{MLO}} \quad (4.7)$$

Where  $X_{MLO}$  represents preprocessed MLO view mammogram images,  $y$  and  $y_{MLO}$  their corresponding labels, and  $N$  is the number of samples. To minimize the loss function, which in this context is Cross-Entropy Loss:

$$\mathcal{L}(W_{MLO}; D_{MLO}) = -\frac{1}{N} \sum_{i=1}^{N_{MLO}} [y_{MLO,i} \log(p(y_{MLO,i} | x_{MLO,i}; W_{MLO})) + (1 - y_{MLO,i}) \log(1 - p(y_{MLO,i} | x_{MLO,i}; W_{MLO}))] \quad (4.8)$$

Based on validation performance, hyperparameters and model weights are adjusted for the best classification accuracy.

$$\hat{y} = F_{final}(F_{CC}(x_{cc}; W_{cc}), F_{MLO}(X_{MLO}; W_{MLO}); W_{final}) \quad (4.9)$$

$F_{CC}$  and  $F_{MLO}$  represent the CC and MLO pathways.  $x_{cc}$  and  $X_{MLO}$  are the input mammograms from each pathway.  $W_{cc}$ ,  $W_{MLO}$  and  $W_{final}$  are the weights for the CC pathway, MLO pathway and final classification system.

Early stopping mechanisms and learning rate scheduling are used to ensure effective convergence. A detailed pseudocode of the EfficientViewNet is given below.

| <b>Pseudocode: EfficientViewNet for Images Classification</b>   |   |
|---|---|
| <b>Input :</b> D_train_CC, D_train_MLO, D_valid_CC, D_valid_MLO |   |
| <b>Initialization:</b>  |   |
| 1: <b>Load</b> CC(x) & MLO(x)                                   | // CC and MLO Pathways                      |
| 2: <b>Define</b> F(x)   | // Fusion for Ensemble Pathway Outputs      |
| 3: <b>Define</b> G(x)   | // Classification Layer for Predictions     |
| 4: <b>Define</b> L(G(x), y)                                     | // Loss Function for Predictions and Labels |
| 5: <b>Initialize</b> weights $w_{CC}$ , $w_{MLO}$               |   |
| 6: <b>Set</b> learning rate $\eta$                              |   |
| 7: <b>Set</b> convergence criteria and early stopping           |   |
| 8: <b>Set</b> epochs $E$  |   |

**Training Loop:**

```

9: for epoch  $\leftarrow$  1 to E do
10:   for each mini-batch (x_batch_CC, y_batch_CC), (x_batch_MLO, y_batch_MLO) in
      Zip(D_train_CC, D_train_MLO) do
11:     // Forward Pass
12:     CC_batch  $\leftarrow$  CC(x_batch_CC)           // CC pathway Predictions
13:     MLO_batch  $\leftarrow$  MLO(x_batch_MLO)       // MLO pathway Predictions
14:     F_batch  $\leftarrow$  F(CC_batch, MLO_batch)    // Ensemble Predictions
15:     G_batch  $\leftarrow$  G(F_batch)               // Final Predictions
16:     // Loss Computation
17:     loss  $\leftarrow$  L(G_batch, y_batch_CC) + L(G_batch, y_batch_MLO) // Same Labels for
      Both Views
18:     // Backward Pass
19:     Compute  $\partial L / \partial w_{CC}$ ,  $\partial L / \partial w_{MLO}$  backward propagation // Gradients of loss w.r.t
      weights
20:     Update w_CC & w_MLO using gradient descent:
21:       w_CC  $\leftarrow$  w_CC - eta  $\times$   $\partial L / \partial w_{CC}$ 
22:       w_MLO  $\leftarrow$  w_MLO - eta  $\times$   $\partial L / \partial w_{MLO}$ 
23:     // Validation
24:     Loss_valid, metric_valid  $\leftarrow$  Validate(P_EVN(x), D_valid_CC, D_valid_MLO)
      // Function to be defined based on metrics of interest
25:     // Convergence Check
26:     if convergence_criteria_met or early_stopping_triggered(Loss_valid, metric_valid)
      then
27:       break
28: return P_EVN(x)
Output: P_EVN(x)

```

**4.3 Evaluation Metrics**

The performance evaluation of the EfficientViewNet model is presented here. These metrics give information about the model's accuracy, precision, recall, and F1 score, allowing for a thorough evaluation of its efficacy. A set of comprehensive evaluation metrics is used to evaluate the effectiveness of the suggested methodology.

**4.3.1. Confusion Matrix**

A confusion matrix is like a table that helps to see how well a model is doing at classifying things. It shows where the model is getting things right and where it is

making mistakes. In breast cancer detection, the model predicts whether a mammogram is normal or abnormal. The confusion matrix for mammogram classification shown in Figure 4. 4.

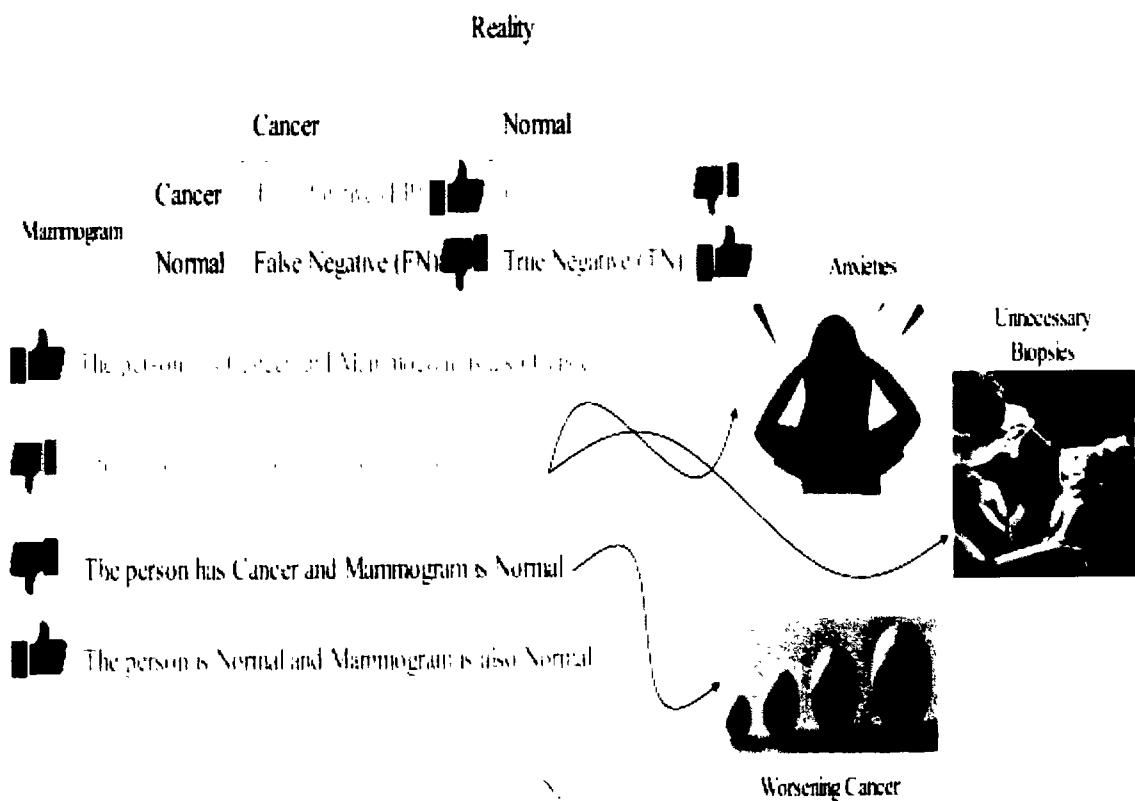


Figure 4. 4 Confusion Matrix

Mammograms correctly identified as showing signs of breast cancer are called True Positives (TP). Mammograms correctly identified as normal without any signs of cancer are termed True Negatives (TN). Similarly, mammograms that are wrongly flagged as showing cancer when they are normal are called False Positives (FP). It is also called a Type I error. Mammograms missed by the model are incorrectly labeled as normal when they do show signs of cancer called False Negatives (FN). Such an error is also called a Type II error.

4.3.2 Accuracy

Out of all the instances, this metric calculates the proportion of correctly classified instances. It offers a comprehensive summary of how well the model forecasts.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4.10)$$

### 4.3.3 Precision

Precision is the measure of the proportion of correct positive predictions out of all positive predictions. Evaluating the precision of positive predictions offers valuable insights into the model's capacity to minimize false positives.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (4.11)$$

### 4.3.4 Recall

The recall measures the ratio of true positive predictions to all actual positives, also called sensitivity or true positive rate. It shows how well the model can find all pertinent instances.

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (4.12)$$

### 4.3.5 F1 measure

The F-measure is calculated as the reciprocal of the arithmetic mean of the reciprocals of recall and precision. It provides a rational viewpoint on the model's trade-off between precision and recall.

$$\text{F1 - Measure} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4.13)$$

## 4.4 Experiments and Results

This section illustrates the model's convergence over multiple epochs. It also provides a comprehensive overview of the EfficientViewNet training process and plots the evolution of various performance metrics.

### 4.4.1 Model Architecture and Training Setup

EfficientViewNet architecture, a member of the EfficientNet family, is used for effective breast cancer detection. It is well known for both its effectiveness and potent feature extraction abilities. A specialized classification head is added, enabling accurate

binary classification for breast cancer detection to further adapt it to the task. The binary cross-entropy loss function was applied next.

Evaluation metrics, such as precision, recall, and accuracy, were selected to evaluate the model's overall performance.

4.4.2 Results and Discussions

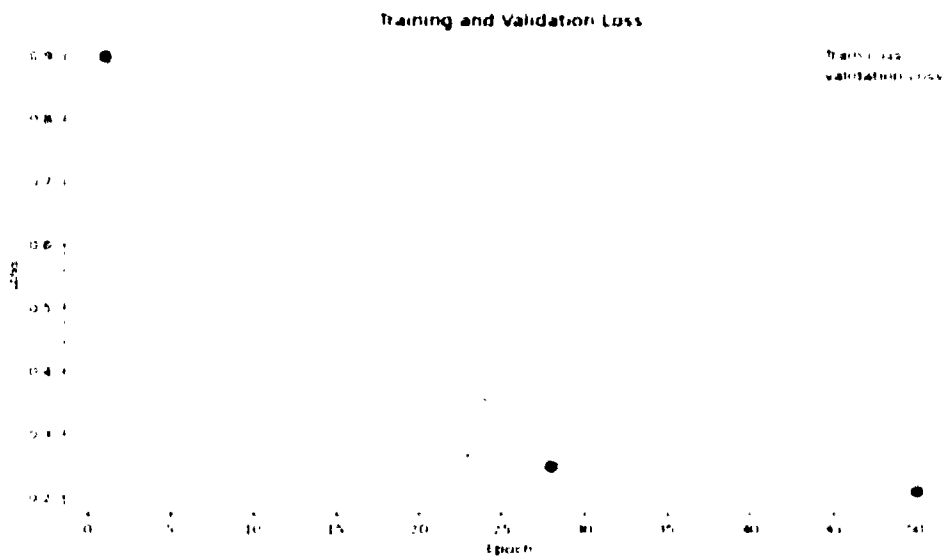
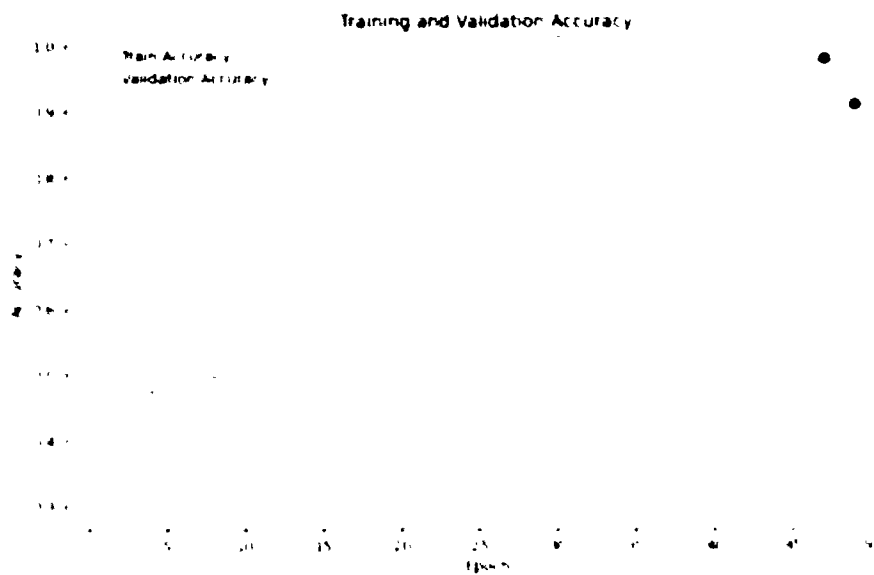


Figure 4.5: Losses of the EfficientViewNet Model

Figure 4.5 illustrates the losses of the EfficientViewNet model for mammogram classification. As the training progresses through the epochs, the model's training loss decreases sharply, signifying a strong initial learning phase where the model rapidly improves its ability to classify the training data. This decrease is expected as the model begins to fit the training data more accurately. Simultaneously, the validation loss decreases, indicating that the model is generalizing well and not just memorizing the training data. The red line starts to go up at one point, which means the model is starting to make more mistakes with new data. The marker on the validation loss at its lowest suggests the most effective model state.



Figure 4.6 shows the accuracy (Training and Validation) of the model. The model's

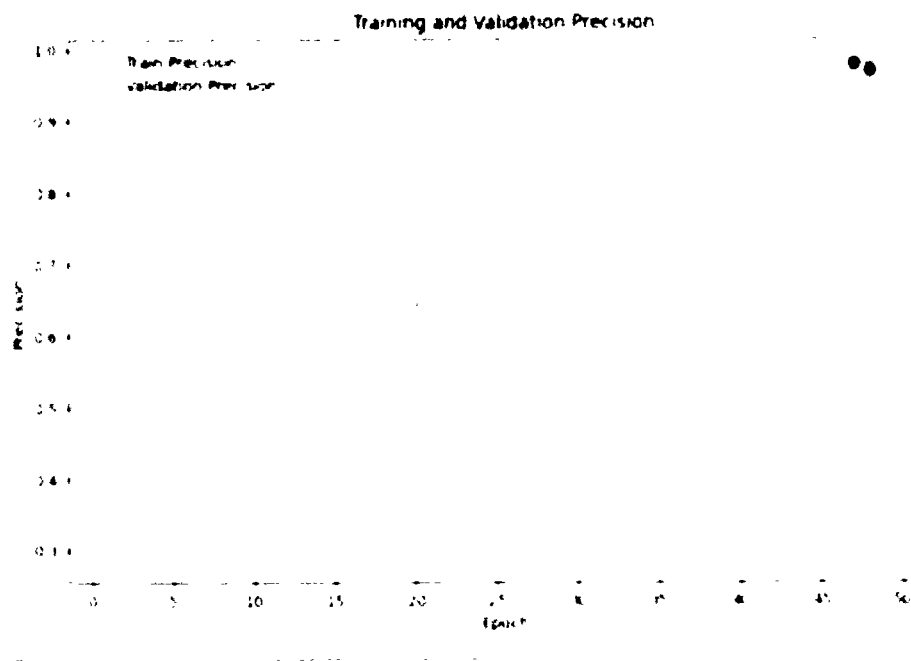


**Figure 4.6: Training and Validation Accuracy of the EfficientViewNet**

training accuracy exhibits an upward trend, indicating that it is learning and improving its ability to correctly classify the training data over time. The progression is smooth, with only minor fluctuations, which suggests that the learning rate and other hyperparameters are well-tuned for the training phase. While trending upward overall, the validation accuracy demonstrates more variability, particularly noticeable in the first half of the training epochs. As the epochs progress beyond the peak validation accuracy point, we observe that the validation accuracy plateaus and decreases slightly, suggesting that the model's generalization ability may not improve further. This suggests that the model's performance on fresh, untested data is less dependent on adding epochs. The training and validation accuracies seem to converge in the last epochs, with the training accuracy leveling off and the validation accuracy showing only slight variations.

As shown in Figure 4.6, the model's precision on the training dataset is low at the outset of training. However, as training progresses, training precision shows a notable upward

trend. This signifies that the model is learning to make fewer false positive predictions on the training data.



**Figure 4.7: Precision Status of the EfficientViewNet Model**

The gradual and consistent increase is indicative of a well-configured learning process. On the other hand, the validation precision exhibits a more variable pattern, especially in the initial epochs. This variability of the model suggests its ability to generalize its learned patterns to unseen validation data fluctuates. Such fluctuations are common in the early stages of training, where the model is still adapting to the broader dataset distribution. In the latter epochs of training, there is a noticeable convergence between the training and validation precision scores. The training precision stabilizes and aligns closely with the validation precision. This convergence signifies that the model's performance on the training and validation datasets is becoming more consistent, indicating successful training.

Figure 4.8 shows that training and validation values increase as the number of epochs grows, indicating an improvement in model performance over time. Training recall shows a steeper increase initially, suggesting that the model quickly learns to identify relevant patterns in the training data. The validation recall also exhibits an upward trend but with a less steep slope than the training recall. This indicates that while the model is learning generalizable patterns, the increase in performance on unseen data is more gradual. The recall values for both sets appear to converge as the epochs approach 50,

with training recall slightly higher than validation recall, which is typical since models are often better at recalling patterns in the data they were trained on.

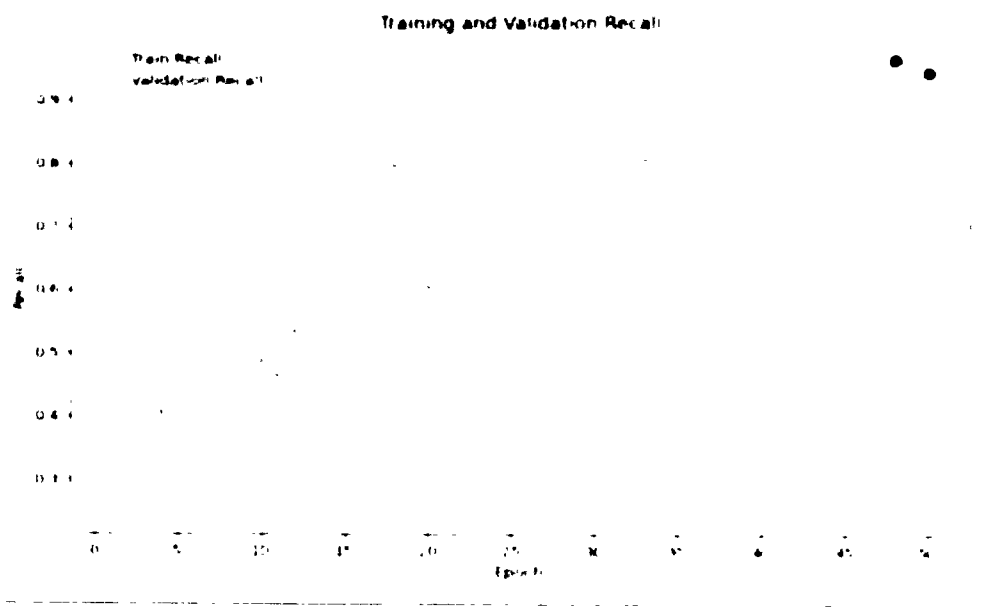


Figure 4.8: Recall of the EfficientViewNet Model

However, the convergence of the lines suggests that the model is not overfitting significantly since the performance on the validation set is close to the training performance.

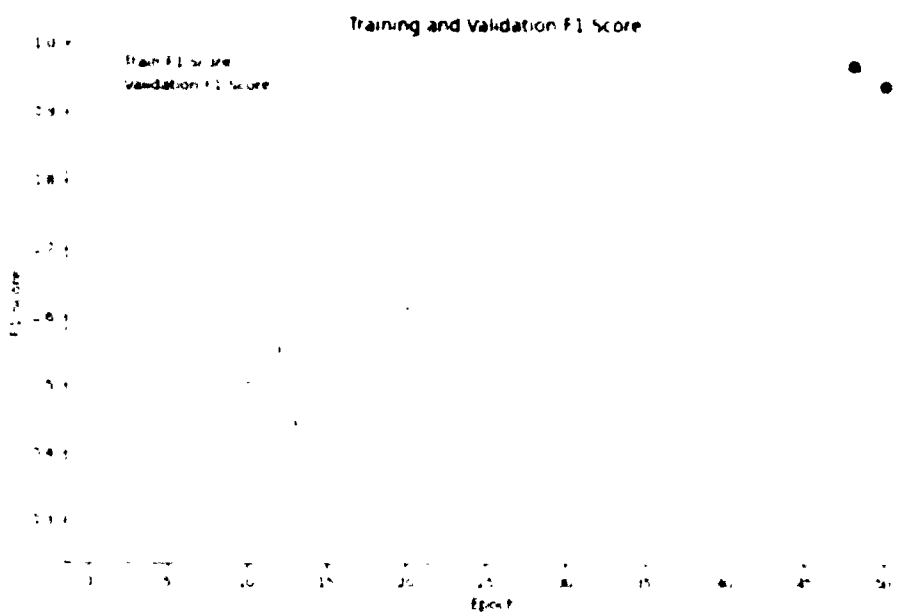


Figure 4.9: F1 Score of the Model

The F1 score in Figure 4.9 exhibits a positive correlation with the number of epochs, indicating that the model is progressively learning and generalizing from the data. The initial F1 score of the training set is low but consistently increases over time, suggesting that the model demonstrates improved accuracy in correctly categorizing the training data. There is a discernible decline in the validation F1 score during the intermediate epochs. This implies that the model's ability to apply its knowledge to new data was not strong, indicating that it had become too specialized for the training data at that specific point. However, following this decline, the validation score rebounds due to modifications made to the model parameters or the implementation of regularization or learning rate adaptation techniques, which aid in restoring the model's ability to generalize. In the later epochs, the training and validation F1 scores reach a stable point, with the training score remaining consistently high. This plateau may indicate that the model has nearly reached its capacity to learn from the training data without further overfitting. The validation score shows slight fluctuations but remains stable, typically a positive sign of the model's robustness.

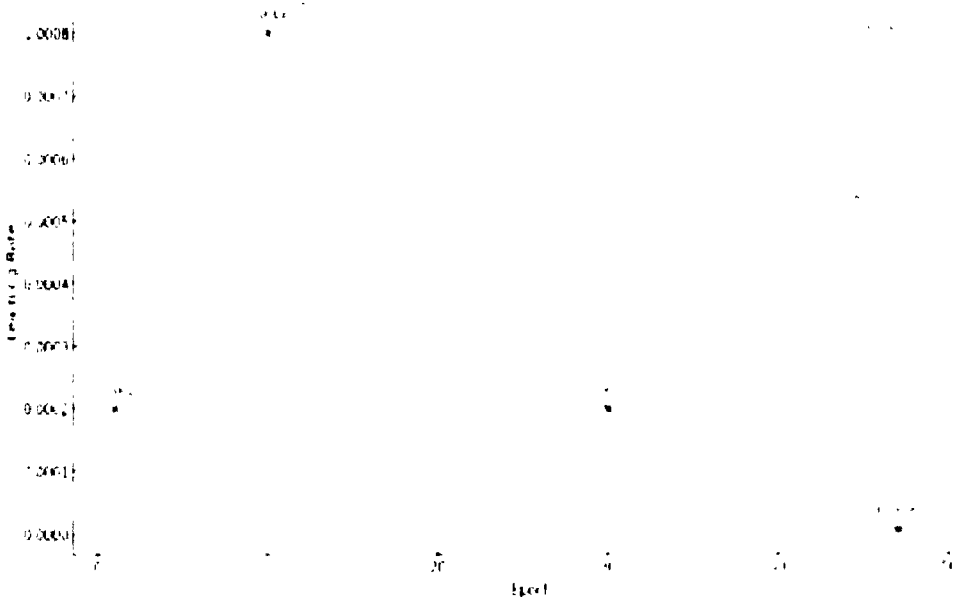


Figure 4.10: Optimization of Learning Rate in EfficientViewNet

Within our novel EfficientViewNet architecture, a dual-pathway approach has been employed to optimize the learning rate for individual views of mammographic images, as shown in Figure 4.10. This distinct strategy acknowledges the heterogeneous nature of mammographic data, allowing for a more nuanced and view-specific learning process. The intricacies of the learning rate optimization are rooted in the architecture's

ability to adjust its parameters adaptively in response to the unique characteristics of each mammogram view. In the context of mammography, the CC and MLO views possess distinct tissue densities and presentations of glandular structures. Our model leverages this disparity by independently calibrating the learning rates, thereby harmonizing the convergence pace of the network with the complexity of the data presented. In the empirical results, we observed that this tailored learning rate approach contributed to a remarkable improvement in classification performance. The model demonstrated an enhanced sensitivity in detecting subtle indicators of pathology, which are often view-specific.

The convergence graphs delineate a clear trajectory of the learning process for each view, highlighting a sharp ascent to the peak learning rate followed by a gradual, controlled descent. This modulation ensured that our model capitalized on rapid learning during the initial epochs, which was crucial for grasping the foundational patterns within the data. Then, it transitioned to a more meticulous fine-tuning phase. By epoch 30, the learning rate approached a near-ideal value, allowing for precise refinement of model weights, culminating in a stable plateau indicative of optimal convergence.

The success of this methodology is evident in the quantitative metrics and the qualitative analysis of the model's performance. EfficientViewNet's adeptness at navigating the learning landscape reaffirms the virtue of a view-specific learning rate strategy.

4.4.3 Comparative Analysis

According to the analysis, the EfficientViewNet model demonstrates superior performance. The threshold guarantees that the model's classification decisions achieve a balance between minimizing false positives and false negatives, which is crucial in breast cancer detection. EfficientViewNet performance is compared with state-of-the-art models such as ResNet, Ensemble, and CNN-GRU across F1 Score, Precision, and Recall. EfficientViewNet shows competitive performance, leading in F1 Score with 0.97, followed by a Precision of .98 and a Recall score of .96

Table 4.3: Performance Comparison with State-of-the-Art Techniques

| Technique        | F1 Score | Precision | Recall |
|------------------|----------|-----------|--------|
| EfficientViewNet | 0.97     | 0.98      | 0.96   |
| DenseNet         | 0.92     | 0.92      | 0.93   |
| ResNet           | 0.90     | 0.91      | 0.92   |
| Ensemble [70]    | 0.76     | 0.86      | 0.76   |
| CNN-GRU [103]    | 0.88     | 0.86      | 0.86   |

The comprehensive performance analysis against existing techniques firmly establishes the EfficientViewNet approach as a breakthrough in breast cancer detection, as shown in Table 4.3 and Figure 4.11 Its exceptional performance across key metrics reinforces its potential for practical implementation in real-world clinical settings, holding the promise to augment diagnostic accuracy significantly.

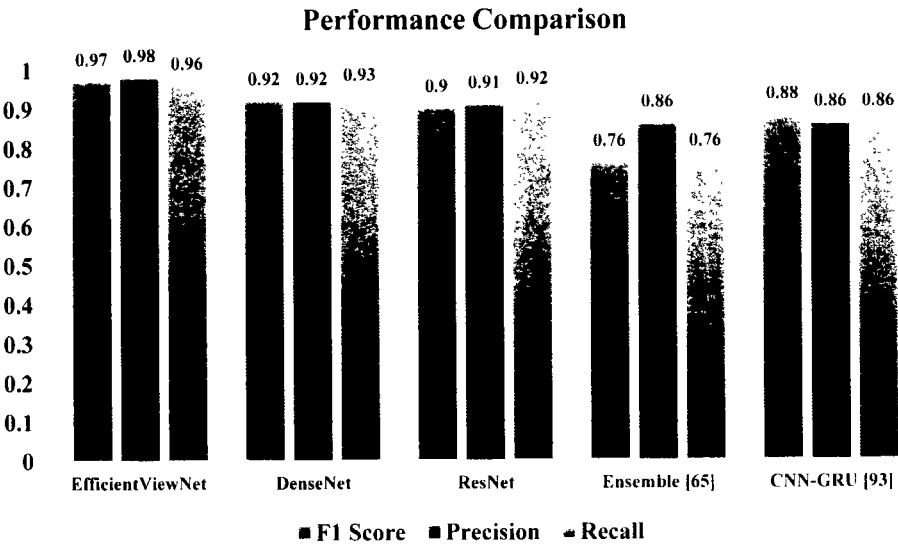
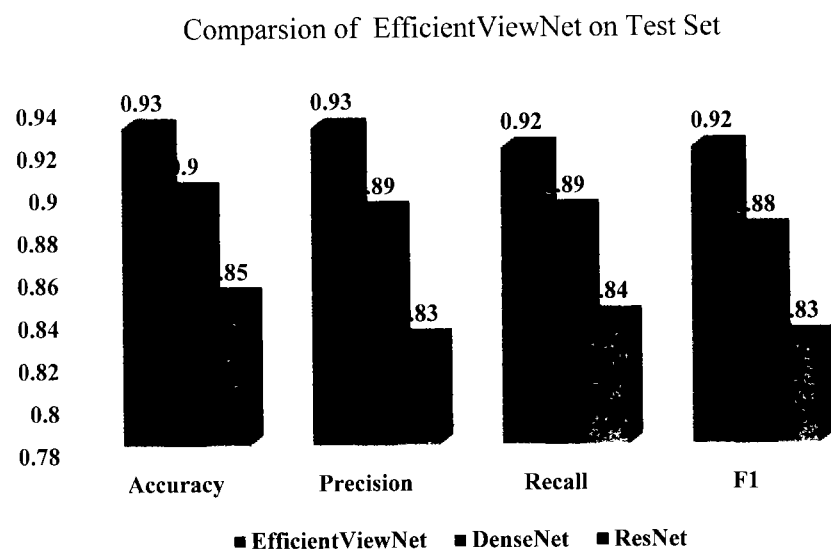


Figure 4.11: EfficientViewNet Vs. Existing Models

Figure 4. 12 compares the performance of EfficientViewNet, DenseNet, and ResNet on the test set. EfficientViewNet outperforms the other models, achieving an accuracy and precision of 0.93, a recall of 0.92, and an F1 score of 0.92. DenseNet ranks second with slightly lower scores of 0.9 for accuracy, 0.89 for both precision and recall, and 0.88 for the F1 score. On the other hand, ResNet has the lowest scores among the three, with 0.85 for accuracy, 0.83 for precision, 0.84 for recall, and 0.83 for the F1 score.



**Figure 4. 12: Performance Comparison of EfficientViewNet vs DenseNet & ResNet on Test Set**

4.5 Chapter Summary

This chapter provides an in-depth exploration of the classification of breast images for cancer detection. The chapter begins by addressing the critical need for accurate breast cancer detection and introduces the EfficientViewNet model, specifically designed to improve diagnostic accuracy through detailed feature extraction from mammographic images. The chapter elaborates on the datasets used followed by EfficientViewNet architecture. It then discusses in detail, highlighting its dual pathways for processing CC and MLO mammographic views. These specialized pathways allow the model to capture intricate details from both views, enhancing its ability to classify breast tumors accurately. The results indicate that EfficientViewNet outperforms existing state-of-the-art techniques, achieving superior scores across all metrics. This is evidenced by its impressive F1 score of 0.97, precision of 0.98, and recall of 0.96, underscoring its reliability in distinguishing between benign and malignant lesions, demonstrating its potential for clinical application in real-world settings.

## Chapter 5

### Optimizing Breast Cancer Detection with Ensemble

#### Learning

Breast cancer detection relies heavily on early and accurate diagnosis to improve treatment outcomes, but traditional methods such as mammography have limitations. This chapter presents an ensemble learning approach, combining the strengths of multiple deep learning models like ResNet, DenseNet, EfficientNet, and AlexNet. By leveraging the predictive power of these models, the ensemble method enhances diagnostic accuracy, reducing variability and increasing reliability in breast cancer detection. The results indicate a significant performance improvement, offering a more robust solution for clinical application.

#### 5.1 Introduction to Ensemble Model

Breast cancer remains one of the most common and concerning forms of cancer that affect women worldwide. Early detection is crucial for improving the prognosis and treatment outcomes of patients. Traditionally, mammography has been the primary method for early detection, proving to be effective in reducing breast cancer mortality rates [104]. However, the interpretation of mammograms is highly dependent on the expertise of radiologists and is subject to variability, which can lead to false positives and negatives. The advent of AI and ML has presented new possibilities for enhancing breast cancer detection techniques. CNNs, a category of advanced artificial neural networks, have demonstrated encouraging outcomes in medical image analysis, specifically in the interpretation of mammograms [105]. These networks can learn to identify complex patterns in image data, potentially exceeding human-level performance in some tasks. Recent research has focused on using various CNN architectures, such as EfficientNet, AlexNet, ResNet, and DenseNet, to detect breast cancer in mammograms. These architectures offer different advantages in accuracy, computational efficiency, and capacity to handle vast amounts of data [106]. Furthermore, the concept of ensemble learning, where multiple models are combined



to improve predictive performance, has gained traction. By integrating the strengths of various CNN architectures, the ensemble methods can offer more robust and accurate predictions than any single model [107].

Breast cancer, a prevalent malignancy among women worldwide, poses a significant public health problem. The World Health Organization has identified breast cancer as the predominant form of cancer affecting women, with a significant impact on millions of individuals annually [108]. Early detection is essential to improve prognosis and treatment success, primarily achieved by mammography. Despite its widespread use, the diagnostic precision of mammography depends on the radiologist's expertise and is prone to variability. This variability can lead to misdiagnosis as false positives or false negatives, underscoring the need for more precise and reliable detection methodologies [109]. Incorporating rigorous AI education into medical imaging has brought a significant and revolutionary change in diagnostic methodologies. CNNs, a specific type of DL model, have exhibited exceptional aptitude in comprehending intricate image data. Their capacity to perceive complex patterns in images positions them as a promising tool for enhancing mammographic interpretation, potentially surpassing conventional methods in accuracy and reliability [110].

Explorations of various CNN architectures, including EfficientNet, AlexNet, ResNet, and DenseNet, have been pivotal in medical image analysis. These architectures bring unique strengths, including computational efficiency and accuracy, which makes them well-suited for analyzing mammographic images. Their characteristics provide valuable information about the complex world of image-based breast cancer diagnosis. An emerging focus in medical AI research is the application of ensemble methods that amalgamate predictions from multiple models to enhance predictive performance. In breast cancer detection, ensemble approaches aim to synthesize the strengths of various CNN architectures to achieve a more accurate and robust diagnostic tool than any single model. Implementing such ensemble methods in cancer detection is particularly interesting due to the critical nature of diagnostic accuracy in clinical decision-making [111]. Using medical imaging benchmarks, compare the effectiveness of leading DL models in breast cancer detection. Implement and evaluate an ensemble model that takes advantage of the predictive power of individual classifiers for enhanced accuracy.

Analyze classifier performance across real and synthetic datasets to understand their generalization capabilities.

This research is motivated by the pressing necessity to enhance the detection of breast cancer, a disease that poses a substantial global health challenge. By enabling early diagnosis and treatment, DL can help overcome the medical infrastructure and resource constraints. Moreover, synthetic data sets could provide a scalable remedy for data constraints, enhancing the accessibility and cost-efficiency of advanced diagnostic tools. The main contributions are listed below.

- The study uses established RSNA and DDSM benchmark data sets alongside a new synthetic data set generated by DCGAN, enriching the breadth of data for model training and testing.
- Conducts an in-depth comparative study of four advanced classifiers, AlexNet, DenseNet, ResNet, and EfficientNet, highlighting their individual and collective efficacy in detecting breast cancer. In addition, an ensemble approach amalgamates the predictive capabilities of individual models, achieving superior accuracy and demonstrating the constructive collaboration of combined classifier wisdom.
- Examines the generalization strength of classifiers across real and synthetic datasets, emphasizing the models' robustness and adaptability to varied data sources.
- This suggests that the benefits of extra epochs on the model's performance on fresh, untested data are negligible. Training and validation accuracies seem to converge in the last epochs, with minor fluctuations in the validation accuracy and a leveling off of the former.

## **5.2 Methodology for Ensemble Model**

To enhance breast cancer detection, this study adopts a multifaceted approach centered on applying CNNs, a class of DL algorithms renowned for their prowess in image analysis. We methodically use four different CNN architectures: AlexNet, DenseNet, EfficientNet, and ResNet, each of which brings a unique set of characteristics and strengths to the task of evaluating medical images. These models are subjected to

rigorous training and validation processes. The overall methodology of the study is presented in Figure 5.1.

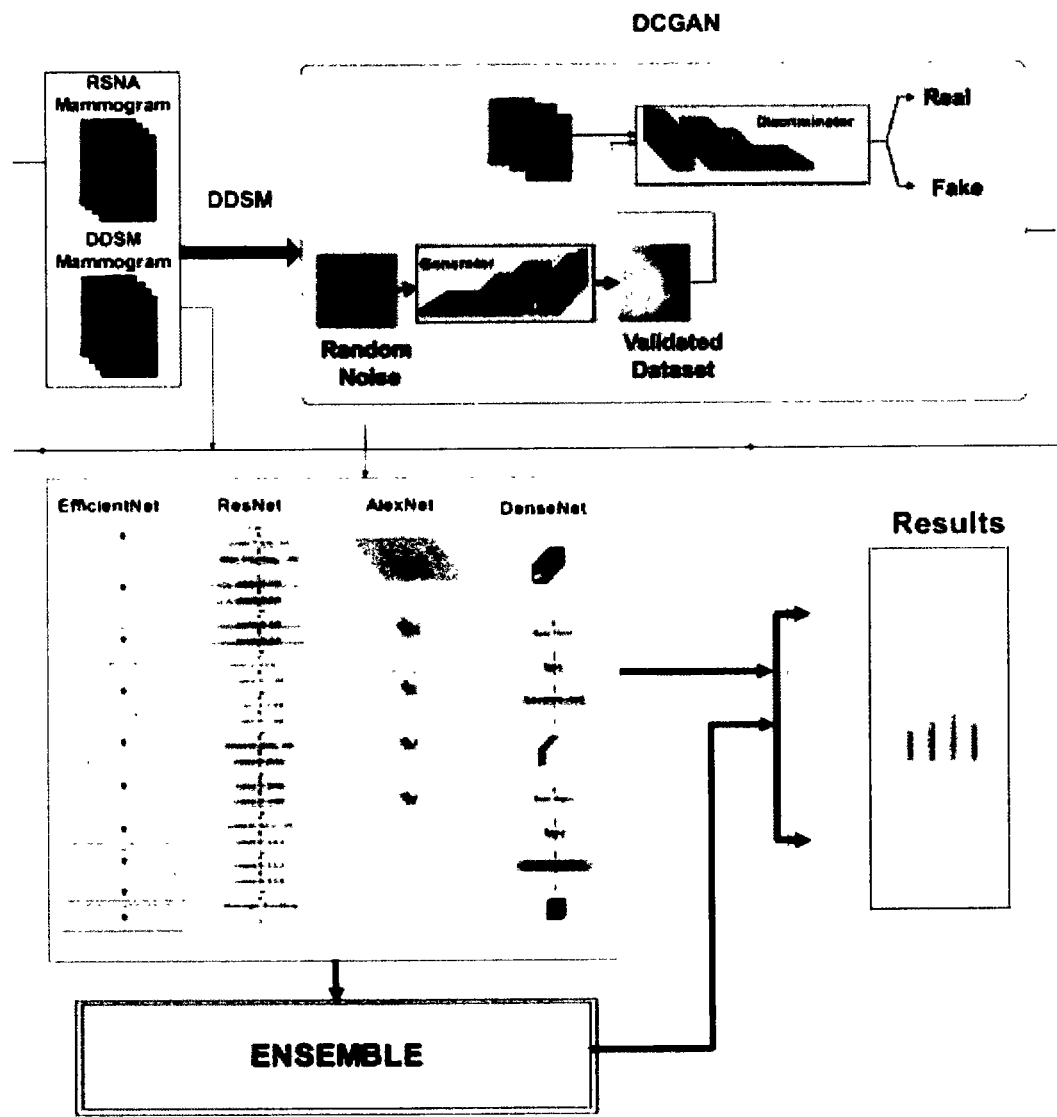


Figure 5.1: Methodology of Ensemble Model

5.2.1 Data collection

The study utilized the Breast Cancer Detection Dataset provided by the Radiological Society of North America (RSNA), which consists of a wide range of mammographic images from a diverse patient population. The RSNA's annual challenge makes it available to the public. It offers a range of annotated cases that have played a crucial role in developing diagnostic models for detecting breast cancer [102]. Complementing the RSNA dataset, the Digital Database for Mammography Screening Mammography (DDSM) [112] was also procured. The DDSM is a well-curated repository that contains thousands of digitized and annotated mammographic images with clinical findings,

which proves invaluable for training and testing ML algorithms in medical imaging. A synthetic data set was generated using DCGAN to mitigate the limitations of data scarcity. This approach involved training DCGANs in DDSM to produce additional high-fidelity mammographic images.

### **5.2.2 Preprocessing and Augmentation of Data**

Data pre-processing ensured that the input data were consistently formatted for CNN models. This included image resizing, cropping, and grayscale normalization. Mammogram images were generated through DCGAN to inflate the dataset.

### **5.2.3 Normalization and standardization of data**

The final step in data preprocessing was normalization and standardization. The pixel values in all images were normalized to a range of  $[0,1]$  to maintain consistency in the intensity distributions. Additionally, data standardization was performed to ensure zero mean and unit variance, thereby accelerating the convergence of DL models during training.

### **5.2.4 Model Selection and Architecture**

When creating a classification system, we carefully chose a group of CNNs, each with a demonstrated history of high performance and unique architectural advancements. The inclusion of EfficientNet is due to its ability to scale network dimensions in a balanced manner, resulting in improved efficiency without sacrificing effectiveness. The DenseNet architecture was selected due to its dense connectivity pattern, which facilitates the reuse of features and improves the flow of information. This results in optimized parameter usage and enhanced model efficiency. ResNet was incorporated into the ensemble to address the issue of vanishing gradients and facilitate the training of deeper network architectures. It achieved this by introducing residual learning and shortcut connections, which preserved the flow of information. Finally, AlexNet was included due to its fundamental contribution to DL in computer vision, providing strong feature extraction capabilities while requiring minimal computational resources.

The collective structure of this ensemble is formulated to exploit the distinct benefits of each model, guaranteeing a comprehensive acquisition of feature representations and a strong ability to generalize. The ensemble is designed to push the boundaries of medical image classification by utilizing the combined strengths of these models.

5.2.5 EfficientNet Configuration

EfficientNet, introduced by Tan and Le [55], is a CNN architecture known for its scalability and efficiency, as shown in Figure 5.2: Basic Architecture of EfficientNet. Balances the network's depth, width, and resolution, tuned through fixed scaling coefficients. In this investigation, an EfficientNet variant with an appropriate scaling factor was chosen to handle the high-resolution nature of medical images, optimizing both accuracy and computational efficiency.

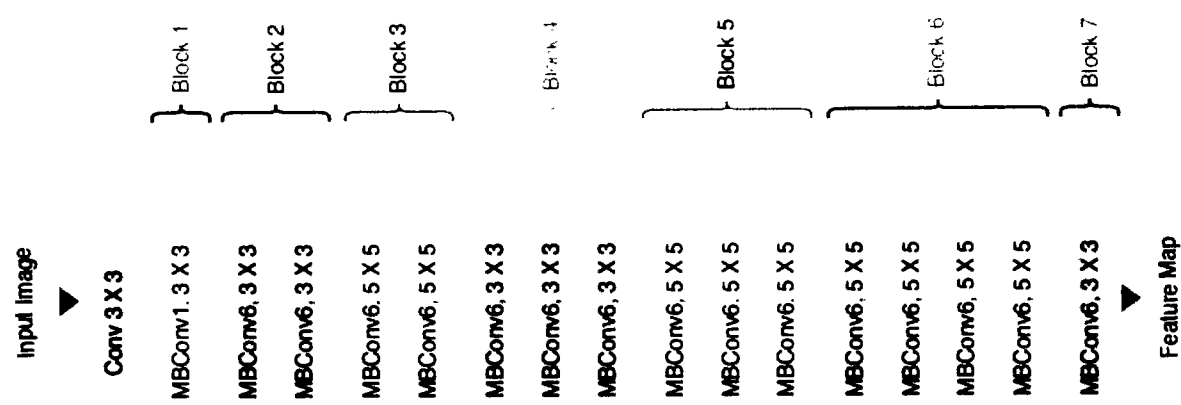


Figure 5.2: Basic Architecture of EfficientNet

5.2.6 AlexNet configuration

AlexNet, developed by Krizhevskii et al. [34], is one of the pioneering CNN architectures, as shown in Figure 5.3. AlexNet comprises 5 convolutional layers in our implementation, followed by 3 fully connected layers. The model uses ReLU activations, dropout, and data augmentation, making it a strong choice for image classification tasks.

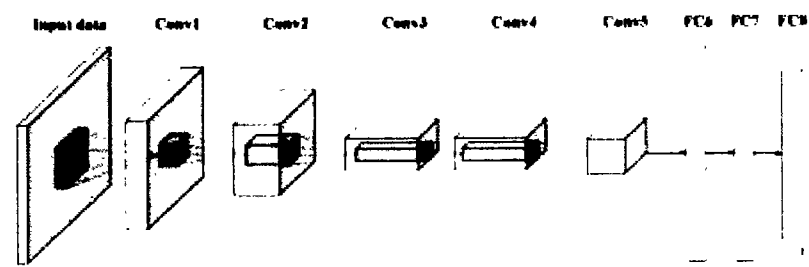


Figure 5.3: Architecture of AlexNet

5.2.7 ResNet Configuration

ResNet, designed by He et al. [113], introduces residual learning to facilitate the training of intense networks. In this research, we use ResNet-50, which contains 50 layers, including residual blocks with skip connections, as given in Figure 5.4. These connections allow it to learn residual functions regarding the layer inputs, enabling the training of deeper networks without suffering from vanishing gradients.

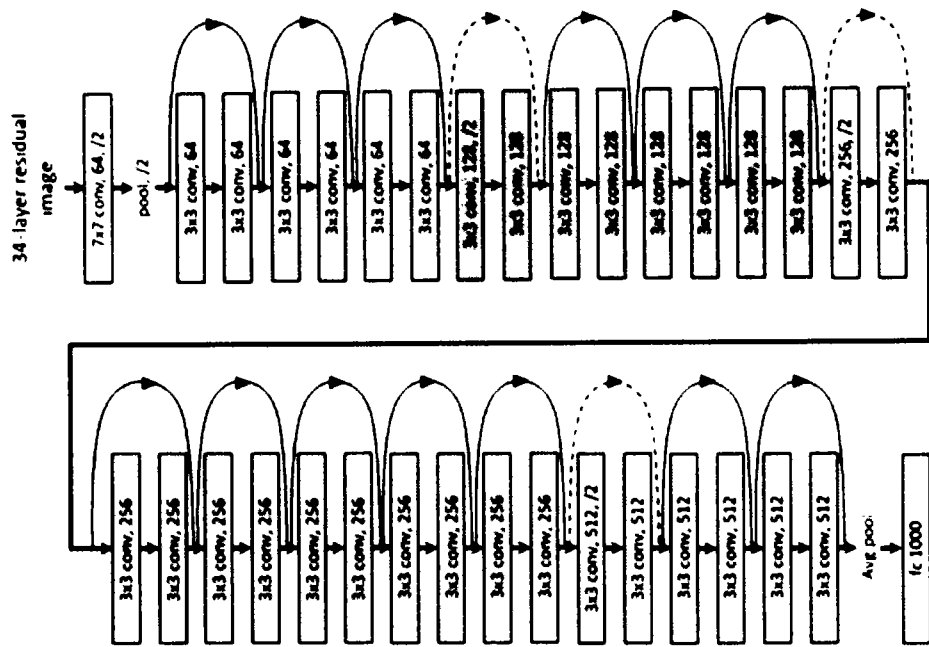


Figure 5.4: Architecture of AlexNet

5.2.8 DenseNet Configuration

DenseNet, proposed by Huang et al.. [114], is characterized by its dense connectivity pattern. We implemented DenseNet-121, which consists of 121 layers, as shown in Figure 5.5. Each layer is connected to every other layer in a feedforward fashion. This architecture, as given in, is particularly efficient concerning computation and parameters, as there is no need to learn redundant feature maps.

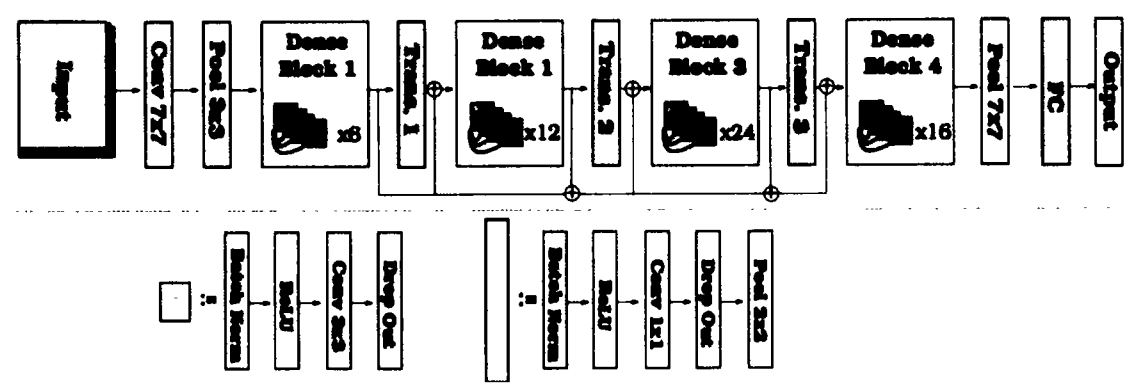


Figure 5.5: DenseNet Architecture

5.2.9 Model Training and Validation

We partitioned the data sets into training, validation, and testing sets for robust training of all models. The RSNA and DDSM data sets were split after an 80-10-10 ratio, whereas the DCGAN-generated synthetic data set was used exclusively for training to enhance feature diversity. We employed a k-fold cross-validation strategy, specifically a 5-fold cross-validation, to ensure our models' generalizability across various data subsets. This approach allowed for a comprehensive evaluation of model performance, minimizing biases associated with random splits of the data.

5.3 Hyperparameter Optimization

This study uses a rigorous hyperparameter optimization process to ensure the optimal performance of each CNN model. This involved using grid search methodology combined with cross-validation to systematically tune the hyperparameters and achieve the best possible results.

Grid search is a widely-used technique for hyperparameter optimization that involves specifying a set of hyperparameters and systematically evaluating all possible combinations within this set. This exhaustive search allows us to explore the hyperparameter space thoroughly and identify the configuration that yields the highest model performance.

To ensure that the hyperparameter tuning process was robust and generalizable, we employed cross-validation. This technique involves partitioning the dataset into several subsets, or "folds," and training the model multiple times, each time using a different fold as the validation set and the remaining folds as the training set. Cross-validation

helps to mitigate overfitting by ensuring that the model's performance is evaluated on multiple subsets of the data, providing a more reliable estimate of its true performance.

The grid search and cross-validation process was iterative. Initial experiments provided a baseline, and subsequent iterations refined the hyperparameter values based on observed performance. Below key hyperparameters that were adjusted. The learning rate controls the step size at each iteration while moving toward a minimum of the loss function. The number of epochs determines how many times the learning algorithm will work through the entire training dataset. Batch size impacts the model's ability to generalize. Smaller batch sizes can lead to noisy gradient estimates, while larger batch sizes might slow down convergence.

The tuning procedure was carefully designed to avoid both underfitting and overfitting. Underfitting occurs when the model is too simple to capture the underlying patterns in the data, while overfitting happens when the model is too complex and captures noise as if it were a true pattern. By iteratively adjusting hyperparameters and using cross-validation, we balanced computational efficiency with model complexity. This balance ensured that our models generalize well to unseen data and perform robustly across different subsets of the dataset.

### 5.3.1 Activation Function

The choice of activation function is crucial for training DL models. In our experiments, we utilized the ReLU as our activation function due to its efficacy in training deep neural networks. The ReLU activation function is given by the following equation.

$$f(x) = \max(0, x) \quad (5.1)$$

This function outputs the input directly if it is positive; otherwise, it outputs zero. ReLU is widely used in DL due to its simplicity and effectiveness in mitigating the vanishing gradient problem, thereby helping the model to improve its accuracy. By allowing only positive values to pass through, ReLU accelerates the convergence of the training process, making the learning of complex patterns more efficient.



### 5.3.2 Training Process and Epoch Management

Each model was subjected to a systematic training process executed over a predefined number of epochs until convergence was observed. Early stopping mechanisms were implemented to terminate training when validation loss ceased to decrease, preventing unnecessary computations and potential overfitting. Model checkpoints were used to save the model's state at the epoch with the lowest validation loss and then for subsequent testing and evaluation.

### 5.3.3 Validation Techniques and Overfit Prevention

In this study, we applied several validation techniques to evaluate the model's performance during training. Real-time monitoring of loss and accuracy metrics in the validation set provided insights into the models' learning progress and generalization ability. To combat overfitting, techniques such as dropout, L2 regularization, and batch normalization were integrated into network architectures. These techniques randomly deactivate neurons and normalize activations, encouraging models to learn more robust features that are not reliant on specific weights, thus enhancing their ability to generalize to new, unseen data.

### 5.3.4 Ensemble model development

An ensemble mechanism is a simple yet powerful method used in this work. Combining the predictions of multiple models is the essence of an ensemble approach to improve overall performance. The fundamental premise is that multiple models each capture distinct facets of the data and that the combined predictions of these models are more dependable and accurate than any one model alone. There are several methods for constructing an ensemble model, such as linear sum, majority voting, and a sophisticated method based on the Dempster-Shafer theory of evidence [115, 116]. These techniques have their pros and cons. The ensemble method in this study is a simple average of the predictions from four distinct models ResNet, DenseNet, AlexNet, and EfficientNet, as given in Figure 5.6. This method involves averaging the predictions of multiple models to produce a final prediction. Simple averaging is easy to implement and computationally efficient, which was essential to keep in view our dataset. Extensive experiments showed that simple averaging delivered competitive and reliable performance across different datasets. The chosen models EfficientNet,

DenseNet, ResNet, and AlexNet individually performed well, and averaging their predictions leveraged their strengths.

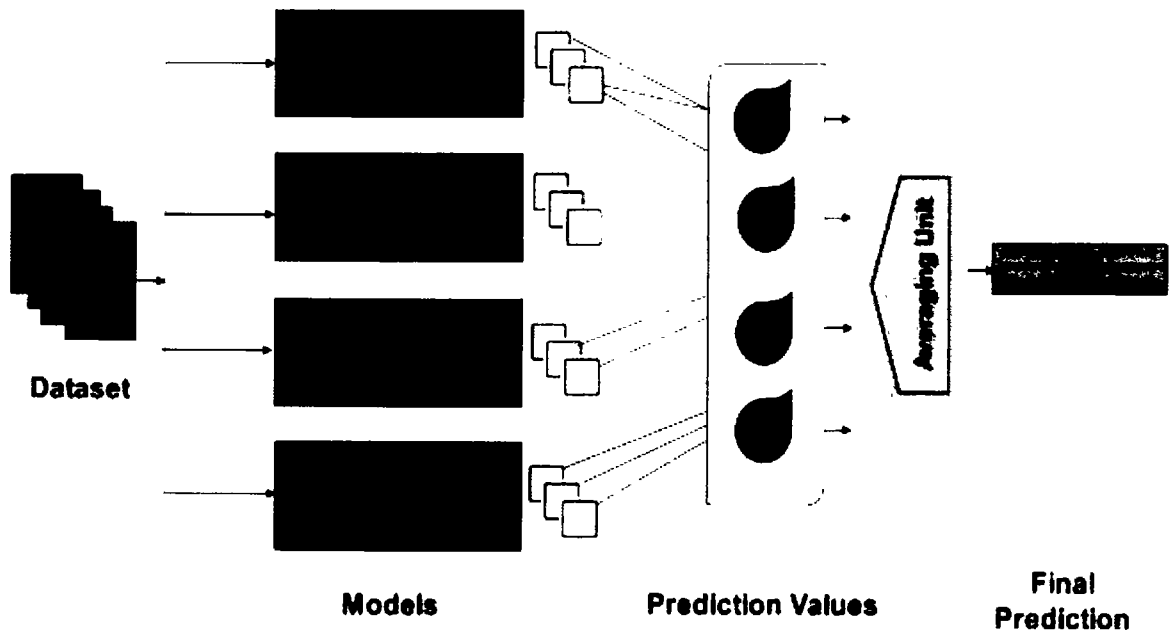


Figure 5.6: Ensemble Layout

This is achieved by summing up the predictions for each model and then dividing them by the number of models. The ensemble can be mathematically represented as

$$\frac{1}{N} \sum_{i=1}^N Pred_i \quad (5.2)$$

where

$N$  is the total number of classifiers in the set.

Where  $Pred_i$  is the prediction made by the classifier in the ensemble. For classification, if the outputs are probabilities, the final class probability for each class  $C$  is:

$$P_{final}(C) = \frac{1}{N} \sum_{i=1}^N P_i(C) \quad (5.3)$$

$P_{final}$  is the direct output. For classification, the class corresponding to the highest  $P_{final}(C)$  is the predicted class. This approach assumes equal weighting for each model, meaning that each model's prediction is considered equally important. The main advantage of this approach is its simplicity and effectiveness in reducing variance and overfitting. By averaging the predictions, it smoothens the individual errors of each

model, potentially leading to more accurate and stable predictions. Details are given in Algorithm1 given below:

---

**Algorithm 1: Mammogram Ensemble Classification**

---

**Input:**

Classifiers  $C \leftarrow C_1, C_2, C_3, C_4$ , each trained on the dataset  $D$

Test instances  $T \leftarrow \{t_1, t_2, \dots, t_n\}$ , where each  $t_i$  is an input vector

**Output:**

Ensemble predictions  $E \leftarrow \{e_1, e_2, \dots, e_n\}$  for test instances  $T$

1: **Procedure** Mammogram\_Ensemble\_Classification( $C, T$ )

2:   **Initialize**  $E \leftarrow []$  to store Ensemble Predictions

3:   **for** each instance  $t_i$  in  $T$

4:     **do**

5:       **Initialize** SumPredictions  $\leftarrow []$        //to store the sum of predictions  
for each class

6:       **for** each classifier  $C_j$  in  $C$

7:         **do**

8:           Obtain Prediction Probabilities  $P_j \leftarrow C_j(t_i)$

9:           SumPredictions  $\leftarrow$  SumPredictions +  $P_j$

10:        **end for**

11:        AvgPredictions  $\leftarrow$  SumPredictions /  $|C|$

12:        PredictedClass  $\leftarrow$  argmax(AvgPredictions)

13:        Append PredictedClass to  $E$

14:     **end for**

15:   **return**  $E$

16: **end Procedure**

---

### 5.3.5 Model training

The ResNet-50 model employed in this study adheres to the conventional architecture comprising 50 layers. The architecture commences with a convolutional layer, succeeded by a MaxPooling layer. Subsequently, it advances through numerous residual blocks, each comprising multiple convolutional and batch normalization layers. The model concludes with a global average pooling layer and a final dense layer (predictions) for classification. The model consists of a total of 23,587,712 parameters, out of which 23,534,592 are trainable and 53,120 are nontrainable.

The implemented model in this study is a sequential model that adheres to the classical architecture of AlexNet. It produces 96 feature maps. After this, a sequence of convolutional layers and max-pooling layers is employed. Subsequently, the model compresses the output and passes it through a sequence of densely connected layers, leading to a final densely connected layer specifically designed for classification purposes. The network comprises three densely connected layers positioned towards

the latter part, with the initial two layers consisting of 4,096 units each, primarily serving the purpose of interpreting features and facilitating the transition to the output layer. The final dense layer is utilized for classification. The model consists of a total of 62,378,344 trainable parameters.

The model utilized in this study is DenseNet-201, which belongs to the DenseNet family renowned for its densely connected convolutional networks. The process begins with an input layer designed to handle images with dimensions of 224x224x3. The model incorporates dense and transition blocks, enhancing its depth and complexity. The architecture is finalized with a global average pooling layer and a dense layer for classification. Following the final global average pool layer, there is a dense layer that contains 1,793,000 parameters. This layer is commonly utilized for classification purposes. The model consists of 27,517,944 parameters, of which 27,258,888 are trainable, and 259,056 are nontrainable.

The ResNet-50 model employed in this study is a deep neural network of 50 layers. It is specifically designed to handle image recognition tasks on a large scale. The architecture commences with an input layer that handles images with dimensions of 224x224x3. Subsequently, a ZeroPadding2D layer, a convolutional layer, and a maximum pooling layer are applied. The central component of the network comprises multiple convolutional blocks, each incorporating batch normalization and activation layers. The network ends with a global average pooling layer and a dense layer. The ultimate dense layer contains 2,049,000 parameters and is specifically tailored for classification. ResNet-50 consists of a total of 23,587,712 parameters, out of which 23,534,592 can be trained.

This study employs EfficientNet due to its high efficiency and effectiveness in processing image data. TensorFlow adheres to these layers. Reapply and transform the images to conform to EfficientNet's input specifications. Subsequently, a dropout layer is implemented to enforce regularization, followed by dense layers for classification.

### **5.3.6 Evaluation Metrics**

The performance evaluation matrices used for Ensemble Model to elevating breast cancer is already presented in Section 4.3. In this study we also utilized AUC for

performance evaluation of the Ensemble model. AUC is particularly valuable in medical imaging and breast cancer detection.

5.4 Results and Discussions

The outcomes of our ensemble classification system, which combines AlexNet, DenseNet, ResNet, and EfficientNet, are covered in this section. The system's effectiveness is rigorously evaluated in breast cancer detection by mammogram classification. This comparative analysis demonstrates our model's enhanced diagnostic accuracy and sheds light on its potential clinical applications in mammography, thereby contributing valuable insights into the field of AI-driven medical imaging.

5.4.1. Performance of Individual Models

The performance of DL models is evaluated with respect to their performance for breast cancer detection. These models include EfficientNet, AlexNet, ResNet, and DenseNet.

Table 5.1: Overall Performance of Models

| Model        | Accuracy (%) | AUC  | Sensitivity (%) | Specificity (%) | F-Measure |
|--------------|--------------|------|-----------------|-----------------|-----------|
| EfficientNet | 93.2         | 94.0 | 92.1            | 93.5            | 92.79     |
| AlexNet      | 87.5         | 88.0 | 85.0            | 88.7            | 86.81     |
| ResNet       | 89.8         | 90.0 | 88.3            | 90.9            | 89.58     |
| DenseNet     | 92.3         | 94.0 | 91.6            | 92.1            | 91.85     |

The results shown in Table 5.1 represent the performance measures of these models. EfficientNet, with an accuracy of 93.2%, AUC of 94%, a sensitivity of 92.1%, and a specificity of 93.5% shows superior performance, indicating its robustness in distinguishing between positive and negative cases effectively. DenseNet closely follows with a similar AUC of 94%, marginally lower accuracy at 92.3%, and slightly less sensitivity and specificity at 91.6% and 92.1%, respectively, suggesting a strong ability to identify true cases with minimal false positives. With a balanced but lower performance across all metrics (accuracy at 89.8%, AUC at 90%, sensitivity at 88.3%, and specificity at 90.9%), ResNet offers a dependable but less optimal solution compared to EfficientNet and DenseNet. AlexNet lags behind the others, with the lowest accuracy of 87.5%, AUC of 88%, sensitivity of 85.0%, and specificity of 88.7%. While it is less precise in distinguishing between conditions, it could still be useful in

scenarios where higher computational efficiency is required and some loss in performance can be tolerated.

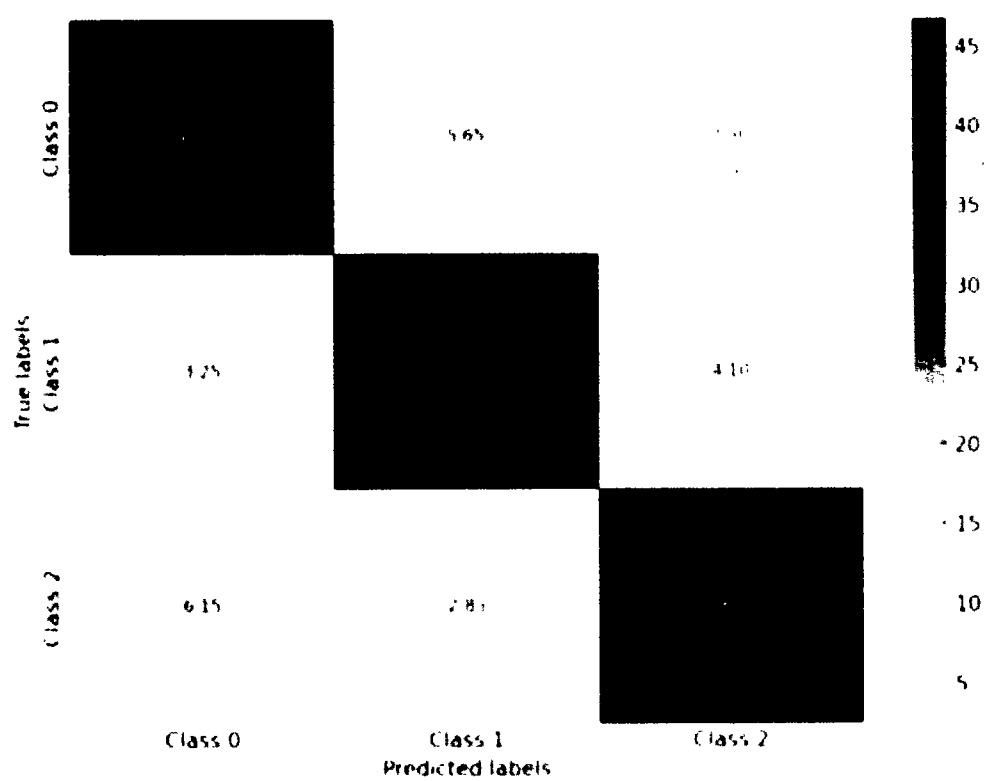


Figure 5.7 Confusion Metrix for EfficientNet Model

Figure 5.7 presents the performance analysis of the EfficientNet model using a confusion matrix. This visual representation effectively highlights the accuracy of EfficientNet in classifying breast cancer images, highlighting its true positives (TP) are those cancer cases predicted as cancer, false positives (FP) are those normal cases predicted as cancer, true negatives (TN) are those cancer predicted as cancer, and false negatives (FN) are those normal cases predicted as cancer.

Figure 5. 8 illustrates the confusion matrix for the ResNet model, providing a detailed breakdown of its performance in identifying breast cancer through image analysis. This matrix effectively highlights the number of correct and incorrect predictions by the model, highlighting its strengths and areas for improvement in distinguishing between cancerous and noncancerous cases.

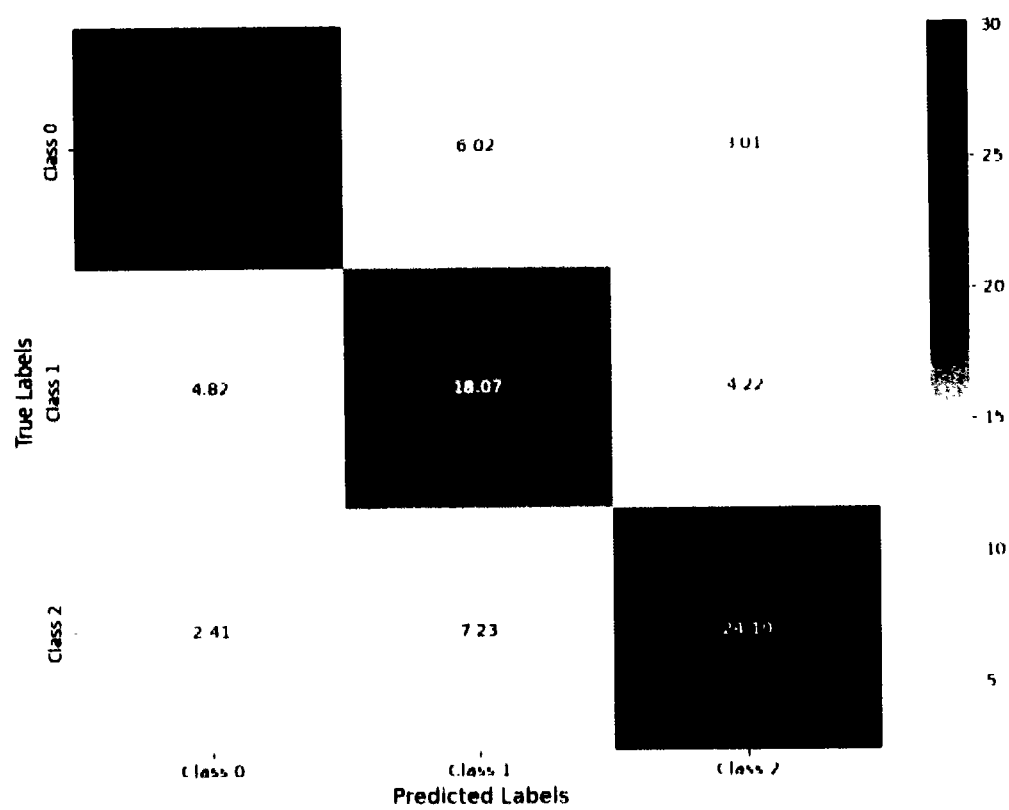


Figure 5. 8: ResNet Model Confusion Matrix

Figure 5. 9 displays the confusion matrix for the AlexNet model, offering a concise visual representation of its classification accuracy in detecting breast cancer.

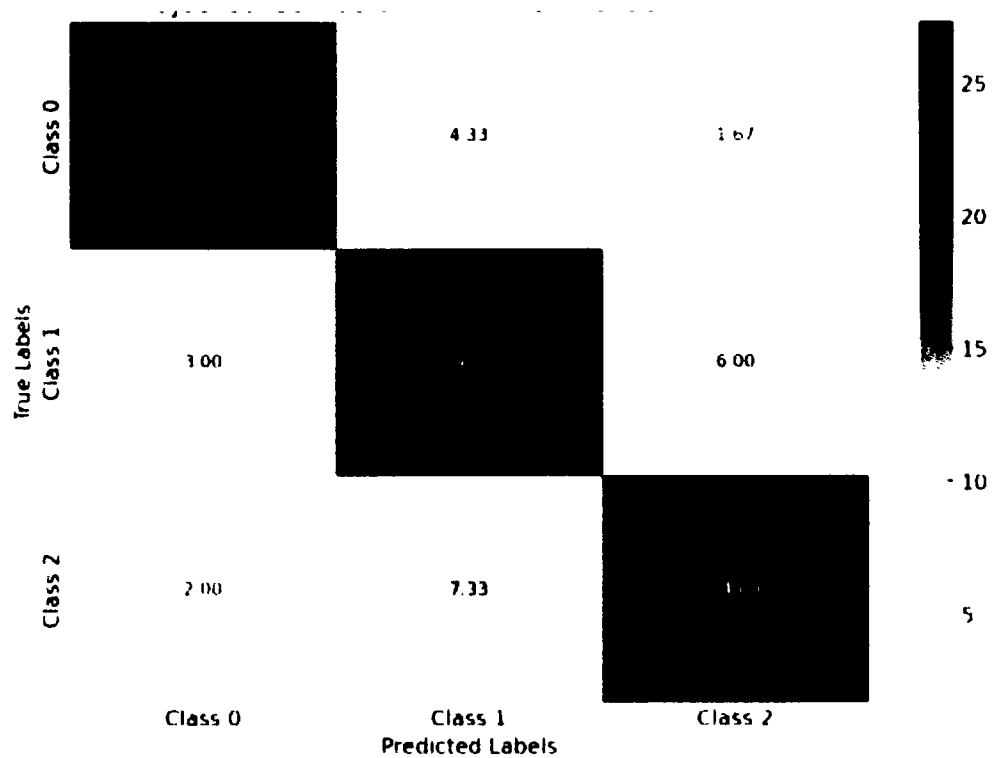


Figure 5. 9: Confusion Matrix for AlexNet model

The matrix details the instances of true and false predictions, both positive and negative, thereby providing a clear insight into the model's diagnostic capabilities and the precision of its image-based cancer detection methodology.

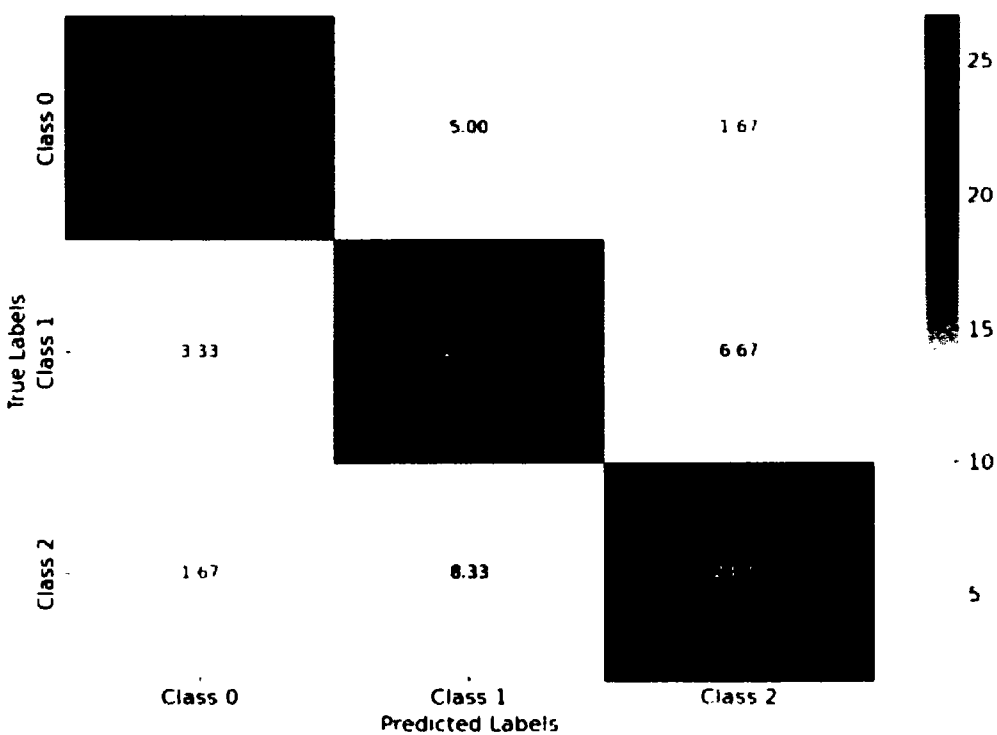
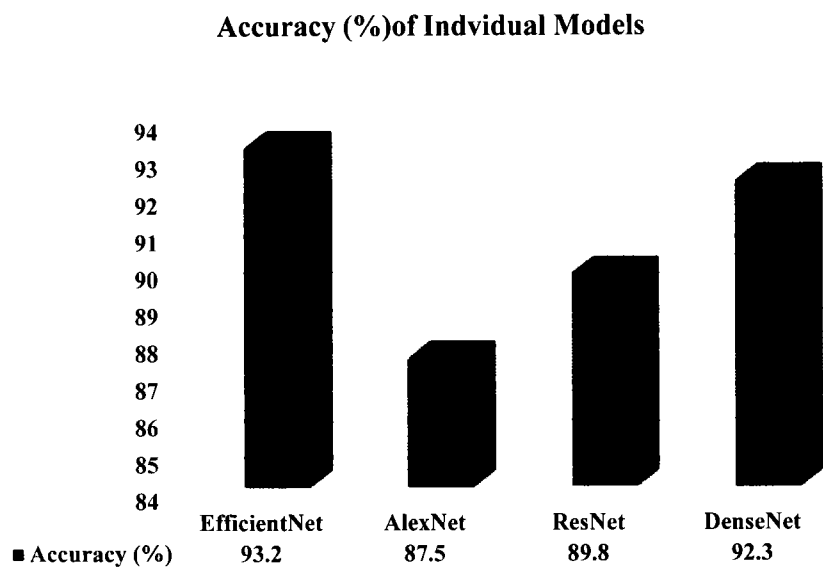


Figure 5. 10: Confusion Matrix for DenseNet model



Figure 5. 10 highlights the confusion matrix for the DenseNet model, which offers an analytical view of its performance in breast cancer detection. This matrix visually depicts the model's accuracy in classifying medical images, highlighting the correct identifications (true positives and negatives) and the misclassifications (false positives and negatives), thereby providing a comprehensive evaluation of DenseNet's effectiveness in medical diagnostics.



**Figure 5.11: Accuracy of Models**

Figure 5.11 provides a comparative overview of the precision of four models applied to mammogram classification. EfficientNet leads with an accuracy of 93.2%, indicating its superior performance in this context. DenseNet follows closely with a precision rate of 92.3%, demonstrating its strong capabilities in feature extraction for medical imaging. With an accuracy of 89.8%, ResNet shows competitive performance, while AlexNet trails by 87.5%, reflecting its older architecture. These results suggest that more modern network designs such as EfficientNet and DenseNet are better suited for the complexities of mammogram analysis in detecting breast cancer.

**5.4.2 Ensemble Approach**

After evaluating the individual performance of EfficientNet, AlexNet, ResNet, and DenseNet in another phase, the models were integrated to form an ensemble: the ensemble method was designed to capitalize on the strengths of each model while mitigating their weaknesses.

Table 5.2: Performance Comparison of Individual Models vs. Ensemble Approach

| Model        | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC  | F-Measure |
|--------------|--------------|-----------------|-----------------|------|-----------|
| EfficientNet | 93.2         | 92.1            | 93.5            | 94.0 | 92.79     |
| AlexNet      | 87.5         | 85.0            | 88.7            | 88.0 | 86.81     |
| ResNet       | 89.8         | 88.3            | 90.9            | 90.0 | 89.58     |
| DenseNet     | 92.3         | 91.6            | 92.1            | 94.0 | 91.85     |
| Ensemble     | 94.6         | 92.4            | 96.1            | 98.0 | 94.21     |

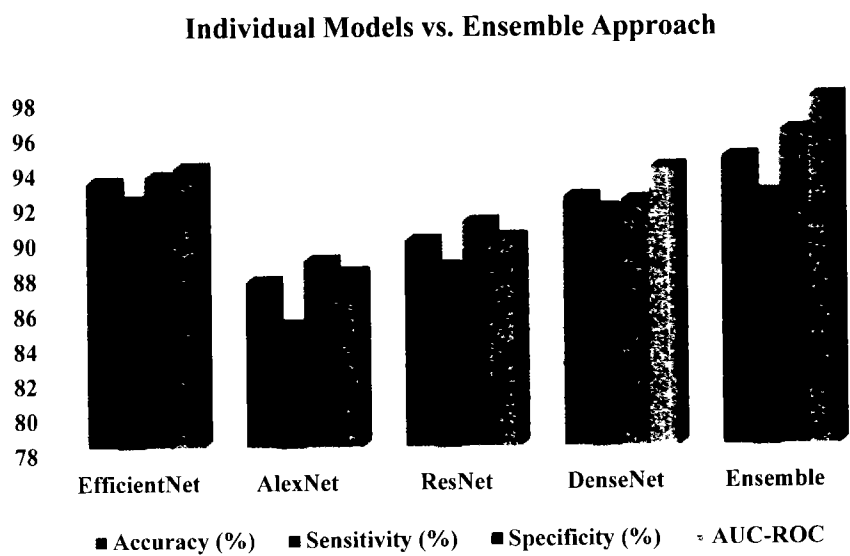


Figure 5. 12: Performance Metrics of ensemble vs. individual models

**Error! Reference source not found.** and Table 5.2 compare individual CNN models EfficientNet, AlexNet, ResNet, and DenseNet and an ensemble approach across four metrics: Accuracy, Sensitivity, Specificity, and AUC. The ensemble method outperforms individual models in all metrics, indicating a synergistic effect that enhances performance. EfficientNet ranks as the most effective single model, particularly regarding accuracy and AUC. AlexNet, while lower in performance, still contributes to the Ensemble's strength. ResNet and DenseNet show solid intermediate performances but are eclipsed by the Ensemble's results, highlighting the advantage of combining models for improved mammogram classification in breast cancer detection. This suggests that the Ensemble approach is highly beneficial in achieving greater diagnostic precision.

Table 5.3: Comparative Analysis of Single Models vs. Ensemble Model

| Criteria        | EfficientNet | AlexNet | ResNet | DenseNet | Ensemble |
|-----------------|--------------|---------|--------|----------|----------|
| Accuracy (%)    | 93.2         | 87.5    | 89.8   | 92.3     | 94.6     |
| Sensitivity (%) | 92.1         | 85.0    | 88.3   | 91.6     | 92.4     |
| Specificity (%) | 93.5         | 88.7    | 90.9   | 92.2     | 96.1     |
| AUC             | 0.94         | .88     | .90    | .94      | 98.0     |

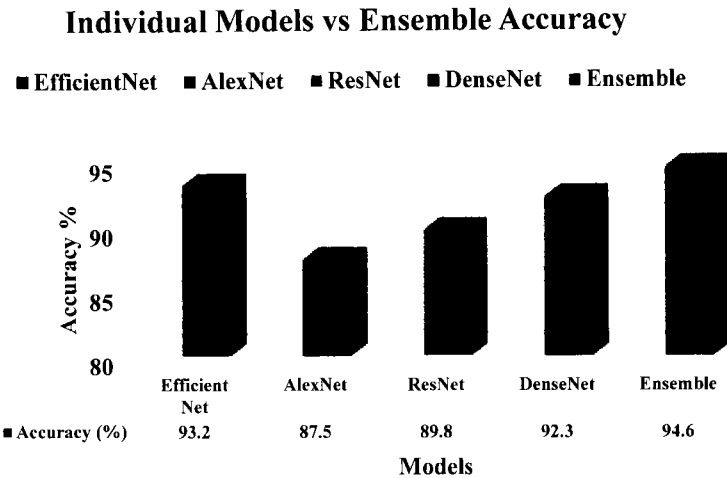


Figure 5.13: Comparative Analysis of Single Models vs. Ensemble Model

Figure 5.13 compares each model's accuracy with the ensemble model's accuracy, emphasizing the ensemble's superior accuracy.

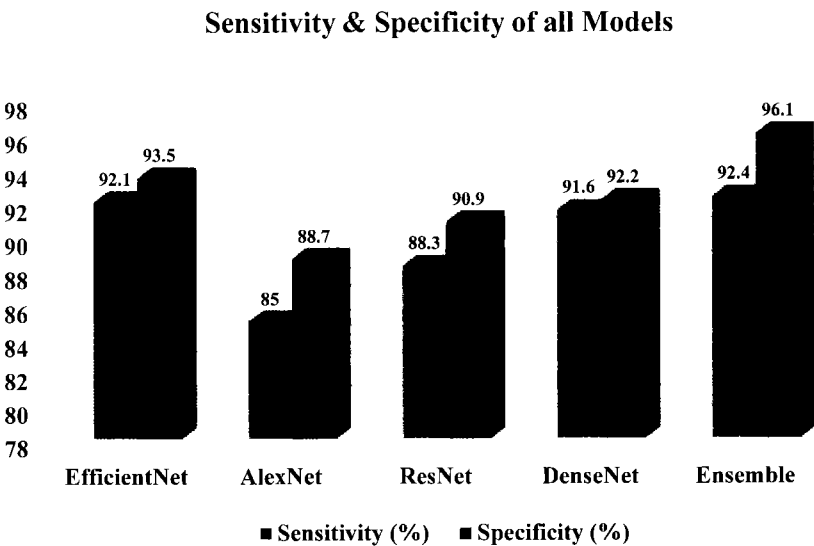


Figure 5.14: Sensitivity and Specificity Comparison

Table 5.3 and Figure 5.14 compare both sensitivity and specificity across all models, illustrating the balanced improvement in the ensemble model.

5.4.3 Comparative Analysis with Existing Models

This section examines how the ensemble approach of integrating EfficientNet, AlexNet, ResNet, and DenseNet compares with existing models in breast cancer screening.

Table 5.4: Performance Comparison with Existing Models

| Model                  | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC  |
|------------------------|--------------|-----------------|-----------------|------|
| Abubacker, et al. [72] | 85           | 86.67           | 83.87%          | 84.0 |
| Arora, et al. [66]     | 88           | 87.0            | 87.2            | 88.0 |
| Mohamed, et al. [71]   | 92.5         | 91.3            | 91.5            | -    |
| Ensemble Approach      | 94.6         | 92.4            | 96.1            | 98.0 |

Table 5.4 summarizes the performance of various CNN models in terms of four metrics: accuracy, sensitivity, specificity, and AUC score. It compares three individual studies by Abubacker et al., Arora et al., and Mohamed et al., with precision ranging from 85% to 92.5%. The table concludes with an "Ensemble Approach" that outperforms individual studies, boasting the highest accuracy of 94.6%, sensitivity of 92.4%, specificity of 96.1%, and an AUC score of 98.0%, indicating a potentially superior model for diagnostic performance.

The ensemble model also excels in specificity (96.1%), suggesting a lower rate of false positives than other methods. The superior AUC value of the ensemble model (0.98) demonstrates its effectiveness in distinguishing between positive and negative cases more accurately than traditional screening methods.

5.5 Chapter Summary

This chapter focuses on the development and evaluation of an ensemble model for breast cancer detection. The ensemble model integrates four high-performing CNN models: ResNet, DenseNet, AlexNet, and EfficientNet. The chapter begins with an overview of ensemble learning and its significance in improving predictive performance. It elaborates on different ensemble techniques; however simple averaging is chosen for its computational efficiency and effectiveness in leveraging the strengths

of individual models. The study uses three primary datasets: the DDSM dataset, synthetic data generated from the DDSM dataset using DCGAN, and a subset of the RSNA breast cancer dataset. These datasets provide a diverse and comprehensive foundation for training and evaluating the ensemble model. A thorough analysis of the ensemble model's performance is presented next. The results highlight the model's robustness and ability to generalize across different datasets. The performance of the ensemble model is rigorously analyzed using metrics such as accuracy, precision, recall, and F1-score. Results demonstrate substantial improvements in predictive performance compared to individual models. The chapter provides strong empirical evidence supporting the use of ensemble methods to enhance predictive accuracy and reliability, contributing valuable insights to the field of medical image analysis.

## Chapter 6

### Lesion Visualization

In breast cancer imaging, medical image visualization is essential because it gives radiologists useful information for diagnosis and treatment planning. Improving the diagnostic precision of DL models depends on visualizing medical images for the ease of domain experts. The use of class activation maps and their significance for breast cancer imaging are covered in this chapter.

#### 6.1 Introduction

Class Activation Map (CAM) is a method that helps make CNNs, which are sometimes viewed as "black boxes," easier to understand. During CNNs' decision-making process, it enables the visualization of areas within an image that are important for identifying a specific disease class. This visualization is achieved by producing heatmaps highlighting the discriminative parts of the image used by CNN to classify it into a specific category. The basic concept of CAM was introduced to provide insight into which features of an image were most influential in the classification decision of a CNN model. The technique involves using the Global Average Pooling (GAP) layers of CNNs to identify the spatial distribution of features that activate for a particular output class. By overlaying the activation maps on the original images, researchers and domain experts can see which pixels contributed most to the model's prediction. This enhances the model's interpretability and helps verify the model's focus on relevant patterns within the image [117].

##### 6.1.1 Importance of CAMs in Medical Imaging

The ability to localize and visualize critical features is important in medical imaging. CAMs can enhance the trust and reliability of automated diagnostic systems by allowing medical professionals to understand the basis of algorithmic predictions [118]. Its application extends beyond validation to the potential discovery of new diagnostic features that may not have been previously recognized. By revealing relevant patterns and features that correlate strongly with certain diseases, these maps can advance

medical knowledge and develop more targeted treatments and better patient outcomes [119]. Integrating these maps into clinical practice supports the move towards precision medicine, where treatment can be tailored based on individual patient profiles. As such, it improves diagnostic model performance and bridges the gap between ML and clinical application, ensuring that the technology truly augments the medical decision-making

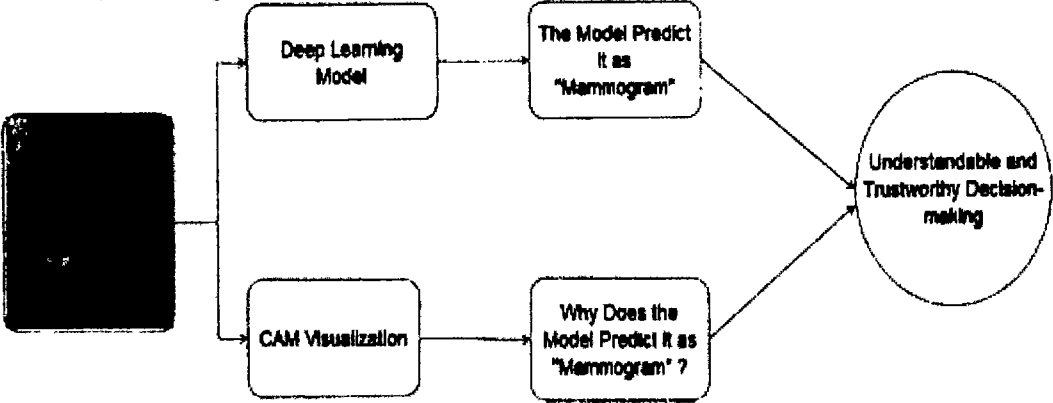


Figure 6.1: An Overview of the CAM Framework

process [120]. A general framework of the CAM visualization is given in Figure 6.1.

6.1.2 Objectives of CAM in Lesions Analysis

The primary objectives of utilizing these maps in the analysis of lesions are multifaceted, aiming to enhance the accuracy and interpretability of mammographic images. These objectives are as follows:

One of the main objectives of CAM in lesion analysis is to improve the diagnostic accuracy of mammography. Highlighting the areas within an image that are pivotal for disease classification assists radiologists in focusing on the regions of interest that may contain pathological features[121]. This targeted approach can lead to more accurate diagnoses, as it reduces the likelihood of overlooking subtle yet critical abnormalities within the breast tissues of a mammogram. Because it can visualize changes in tissue structure that may not be easily visible to the human eye, it can be extremely helpful in the early detection of diseases. Diseases that are difficult to detect in their early stages, like some forms of breast cancer, can appear as slight changes in tissue.

Identifying the exact location and extent of tissue defects enables healthcare providers to tailor treatments to patients' needs, potentially leading to more targeted therapies and better disease management.

It also serves as a research tool that can uncover new insights into the characteristics of various lesions. Researchers and other stakeholders can use it to identify patterns and markers indicative of specific diseases, contributing to the broader understanding of disease mechanisms and progression. CAM's interpretability helps bridge the gap between AI models and their clinical application. Offering a visual explanation for AI-driven diagnostic suggestions fosters trust among medical practitioners in using AI tools, facilitating their adoption in clinical settings.

It can be utilized as an educational tool for medical students and trainees. Visualizations make it easier to understand the features most indicative of certain pathologies, enhancing the educational process and providing a solid foundation for future clinical practice. It can also act as a quality control mechanism. Verifying that the AI model's focus aligns with the clinically relevant features of an image ensures that the model is not basing its decisions on artifacts or irrelevant aspects of the data, thereby validating the model's performance and reliability.

## 6.2 Theoretical Background

DL offers sophisticated algorithms that can learn from and interpret complex data; it has completely changed the field of medical diagnosis. These algorithms, particularly neural networks, can automatically detect patterns and characteristics in medical images indicative of various health conditions. The application of DL in medical diagnosis is diverse, ranging from detecting malignancies in medical images [24]. The strength of DL lies in its capacity to handle vast amounts of data and learn features without the need for explicit programming. This is particularly beneficial in medical diagnosis, where the nuances of medical images can be subtle, and the volume of data for analysis is significant. DL models, through their multilayered structure, can extract hierarchical features, with each layer building on the previous one to refine the understanding of the data. This process allows for capturing intricate patterns that may not be visible to the human eye, leading to more accurate and early diagnosis of diseases [122].



### 6.2.1 CNN for Image Recognition

CNNs are a subclass of deep neural networks that are now essential for image processing and recognition tasks. They are specifically designed to process pixel data and are adept at handling images. CNNs use convolutional layers that apply filters to the input image to create feature maps, highlighting important features such as edges, textures, and shapes. These features are then pooled and flattened to be passed through fully connected layers, resulting in the ability to classify the image into various categories [123]. In medical imaging, CNNs have been employed to identify and classify various diseases from imaging modalities like X-rays, mammograms, MRIs, and CT scans. The ability of CNNs to learn from large datasets of annotated images allows them to recognize complex patterns associated with specific medical conditions [17].

### 6.2.2 Feature Activation in CNNs

Feature activation in CNNs refers to how specific neurons in the network become activated in response to certain features or patterns within the input data. When a CNN processes an image, different layers capture different aspects of the image. Early layers may detect simple features like edges and corners. In contrast, deeper layers can identify more complex patterns corresponding to higher-level concepts, such as textures or objects within the image. Activating these features is critical to how CNNs make sense of visual information. By examining the activations, researchers and practitioners can gain insights into what the network is "seeing" and which features are most influential in decision-making. This understanding is crucial for refining the model's performance, as it can indicate whether the network is focusing on relevant patterns or being misled by noise or irrelevant information in the data. Activations are not just binary; they occur at varying intensities, which allows the network to express a degree of certainty about what it perceives in the input data. This gradation is part of what gives CNNs their robustness in handling variations within the data. By backpropagating errors based on these activations, CNNs learn to adjust their weights, refining the feature detection process for future inputs.

6.2.3 Evolution of CAM Techniques

These techniques have evolved to address the need for more interpretable DL models, especially in domains where understanding model reasoning is as important as the output. Its original method was proposed to directly visualize which parts of an image are relevant to a CNN's classification decision by using the weights of the last fully connected layers to produce a heatmap [117]. Since then, several variations and improvements of CAM have been developed. Grad-Gradient-weighted (Grad-CAM) is a technique that uses the gradients flowing into the last convolutional layer of the CNN to understand the importance of each neuron for a decision of interest. This method allows for coarse localization maps highlighting important image regions to predict the concept [124]. Further advancements have led to techniques like Grad-CAM++, which provide better visual explanations for CNN-based models, especially in cases with multiple occurrences of the same class in the image. These methods continue to refine the granularity and accuracy of the visual explanations provided by these techniques, making them more useful for practical applications and increasing trust in automated systems [125].

6.3 Results and Discussion

The results of using these maps to recognize and visualize breast lesions on a series of mammograms are shown in the following section. The resulting heatmaps were then examined.

Table 6. 1: Summary of CAM Analysis Results

| Image ID | Predicted Pathology | CAM Highlighted    | Clinician Validation | Remarks   |
|----------|---------------------|--------------------|----------------------|---|
| IMG001   | Benign              | Tumor Periphery    | Confirmed            | CAM correctly identified the periphery, which is often indicative of benign tumors. |
| IMG002   | Malignant           | Central Mass       | Confirmed            | Central mass enhancement is characteristic of malignancy.                           |
| IMG003   | Benign              | Diffuse Patterning | Partial Match        | CAM highlighted diffuse areas, but the clinician notes                              |

| Image ID | Predicted Pathology | CAM Highlighted  | Clinician Validation | Remarks  |
|----------|---------------------|------------------|----------------------|--|
|          |                     |                  |                      | some of the regions of interest were not captured.                 |
| IMG004   | Benign              | Irregular Shapes | Confirmed            | The irregular shape of activation aligns with the tissue patterns. |
| IMG005   | Benign              | Swollen Area     | Confirmed            | The swelling was detected, as highlighted by CAM                   |

Table 6. 1 summarizes the CAM results, correlating the predicted pathology with the specific regions highlighted by this technique and subsequent clinician validation.



Figure 6.2 Benign Tumor Detection

Figure 6.2 illustrates a CAM heatmap overlaid on a medical image of a benign tumor. The highlighted periphery is consistent with the benign nature of the tumor, as confirmed by clinical diagnosis.

A team of clinicians evaluated the heatmaps to validate the accuracy of the model's predictions. The validation process involved comparing the CAM-highlighted regions

with the actual locations of lesions as determined by radiological analysis. The concordance rate between the machine’s predictions and clinical findings was calculated to assess the model's performance. The visualizations produced allow for a more transparent assessment of the model's decision-making process, essential for clinical acceptance and further application in existing healthcare settings.

6.3.1 Quantitative Analysis

The quantitative analysis focuses on the statistical evaluation of the results obtained. The performance metrics are based on comparing the CAM-identified regions and the ground truth annotations provided by medical experts. The following table presents some of the mammograms summarizing the performance in detecting various lesions in the mammograms.

Table 6.2: Performance Metrics for CAM Detection

| Metric               | Value | Description  |
|----------------------|-------|--|
| Accuracy             | 92%   | Percentage of total predictions that were correct.                       |
| Precision            | 89%   | Percentage of identifications that were correct.                         |
| F1 Score             | 91.5% | Harmonic means precision and recall.                                     |
| Recall               | 94%   | Percentage of actual positives that were correctly identified.           |
| Area Under ROC Curve | 0.95  | Measure of the ability of the classifier to distinguish between classes. |
| Specificity          | 88%   | Percentage of negatives that were correctly identified.                  |

Table 6.2 provides a comprehensive overview of the performance metrics used to evaluate the effectiveness of the CAM technique in the context of breast lesions in mammograms.

6.3.2 Qualitative Analysis

Qualitative analysis involves a subjective assessment of the heatmaps by a panel of expert clinicians.

Table 6.3: Clinician Assessment of CAM Heatmaps

| Image ID | Clinician | Assessment Score | Comments   |
|----------|-----------|------------------|--|
| IMG001   | A         | 4.5/5            | Excellent correlation with the pathological features.        |
| IMG002   | B         | 3.0/5            | Good detection, but some areas of interest were missed.      |
| IMG003   | C         | 4.0/5            | Accurate localization, minor discrepancies in the periphery. |
| IMG004   | D         | 2.5/5            | Poor correlation, model misinterpretation.                   |
| IMG005   | E         | 5.0/5            | Perfect alignment with clinical diagnosis.                   |

Table 6.3 summarizes the qualitative assessments provided by clinicians, with scores reflecting the perceived accuracy and relevance of the heatmaps. It helps identify the model's strengths in detecting clinically relevant features and its weaknesses where improvements are necessary. This feedback loop is vital for the iterative refinement of the model to ensure its alignment with clinical expectations and requirements.

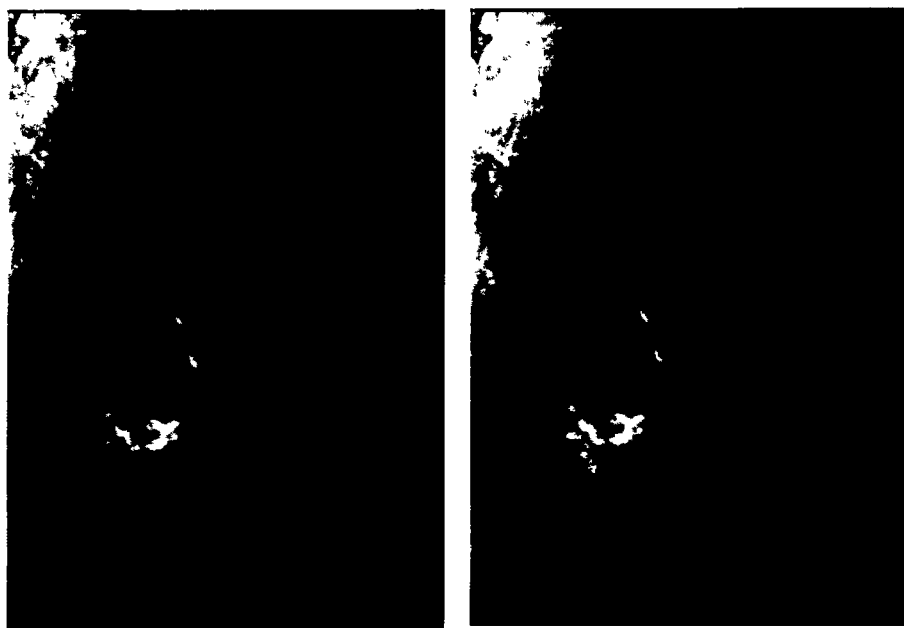
6.3.3 Assessment by Medical Experts

Experts reviewed the heatmaps to verify the correlation between highlighted regions and known pathological features. This section presents a summary of expert evaluations of the results.

Table 6.4: Expert Evaluation of CAM Heatmaps

| Image ID | Expert | Diagnosis      | CAM Concordance | Expert Comments                                    |
|----------|--------|----------------|-----------------|--|
| IMG101   | A      | Lesion         | High            | It aligns well with lesion margins.                |
| IMG102   | B      | Tumor          | Moderate        | It partially overlaps with the tumor area.         |
| IMG103   | C      | Inflammation   | Low             | The highlights do not match inflammation sites.    |
| IMG104   | D      | Healthy Tissue | Very High       | It correctly identifies no defects.                |
| IMG105   | E      | Cyst           | High            | Accurate depiction of cyst location by these maps. |

Table 6.4 shows the evaluations made by different medical experts on a series of images, providing insights into the accuracy of CAM in clinical practice.



**Figure 6.3: Expert Review of CAM Heatmap for Tumor Identification**

Figure 6.3 displays a CAM heatmap used for tumor identification with expert annotations indicating the degree of overlap between the heatmap and the actual tumor boundaries.

## 6.4. Discussion

Understanding the model's focus and the highlighted regions' clinical significance is necessary for interpreting these findings. By generating heatmaps that show the regions of an image that significantly contribute to the model's output, it offers a visual explanation for the choices made by CNNs. The heatmaps can be a link between the two-dimensional input space understandable to humans and the high-dimensional feature space that CNN has learned.

By providing a visual aid to the areas within an image that a CNN finds most indicative of a particular pathology, It can assist clinicians in focusing their examination and potentially uncovering aspects of the idea that may not have been immediately apparent. It serves as a second opinion for radiologists and other medical professionals, offering a data-driven perspective that complements their expertise. The heatmaps can also be used as a teaching tool, helping train medical students and residents by visually demonstrating the features often associated with specific pathologies.

Table 6.5: Clinical Validation of CAMs

| Image ID | Pathology    | CAM Validation | Clinical Outcome | Concordance Rate |
|----------|--------------|----------------|------------------|------------------|
| IMG201   | Lesion       | High           | Confirmed Lesion | 95%              |
| IMG202   | Tumor        | Moderate       | Confirmed Tumor  | 75%              |
| IMG203   | Inflammation | Low            | No Inflammation  | 60%              |
| IMG204   | Healthy      | Very High      | Healthy Tissue   | 99%              |
| IMG205   | Cyst         | High           | Confirmed Cyst   | 90%              |

Table 6.5 summarizes the rate at which CAM heatmaps align with clinical outcomes, providing a quantitative measure of the technology's reliability. The transparency and interpretability of AI decisions are crucial for gaining the trust of both clinicians and patients. As such, integrating CAMs into clinical workflows must be approached with an emphasis on augmenting rather than replacing human judgment, ensuring that the final diagnostic decisions remain within the purview of trained medical professionals.

6.5 Chapter Summary

In normal circumstances, the output of a CNN is difficult to interpret by domain experts. CNNs are often considered "black box" models because they process data through multiple layers of non-linear transformations, making it challenging to understand how the model arrives at its decisions. The complex, high-dimensional nature of the features extracted by CNNs can be opaque, meaning that even though the model may achieve high accuracy, it can be difficult for experts to understand which specific aspects of the input data (such as an image) influenced the model's predictions. This lack of transparency can hinder trust and acceptance in fields like medical diagnostics, where understanding the reasoning behind a decision is crucial. The critical role of lesion visualization in breast cancer imaging is discussed in this chapter, emphasizing the use of CAMs to enhance the interpretability of CNNs. CAMs generate heatmaps that highlight the areas within medical images that are important for disease classification, making it easier for domain experts to understand and trust the model's decisions.

## Chapter 7

### Conclusion and Future Work

#### 7.1 Conclusion

The culmination of this thesis marks a significant milestone in the quest to enhance the diagnostic accuracy of breast cancer through the use of advanced CAD systems. At the core of this endeavor was the challenge posed by the limited availability of annotated mammographic images. To address this, the study employed DCGAN, which proved instrumental in generating a synthetic yet highly realistic dataset of mammographic images. This enriched dataset was carefully validated using innovative similarity assessment techniques, confidence interval analysis and the critical eye of seasoned radiologists, ensuring that the data remained diverse and representative.

To tailor the CAD system to the specific nuances of mammography, the research introduced the Two-pathway EfficientViewNet model. This model was designed with a keen understanding of the distinct characteristics of CC and MLO mammographic views. The architecture's unique dual pathway enabled the extraction of pivotal features from each view, leading to a remarkable improvement in tumor classification accuracy. The empirical results speak for themselves, with the model achieving an F1 score of 97.0%, a precision of 98.0%, and a recall of 96.0%. These figures are not just numbers but represent a substantial leap over the performance benchmarks set by previous models.

The exploration into the realm of ensemble learning revealed another layer of potential. By amalgamating the strengths of various established models such as EfficientNet, DenseNet, ResNet, and AlexNet, the study ventured beyond the capabilities of individual classifiers. The resultant ensemble model manifested a robustness reflected in a precision of 94.6%, a sensitivity of 92.4%, a specificity of 96.1%, and an AUC of 98.0%. Notably, EfficientNet emerged as a standout among the individual models, delivering solid performance metrics on their own. A crucial addition to the CAD system's toolkit was the integration of CAMs. This technique illuminated the decision-making process of the CAD systems, providing visual insights into the areas of interest



within mammographic images that influenced the model's predictions. Such clarity enhanced the interpretability of the results and served to bridge the oft-cited gap between the complexity of AI models and their practical clinical applicability.

## 7.2 Future Work

In the future, we aim to explore the realm of GANs more deeply by investigating state-of-the-art architectures that promise enhanced image quality and refined control in image generation. One of our focal points will be to amplify the image generation process by integrating GANs with other sophisticated ML techniques, such as autoencoders.

A critical aspect of our forthcoming work will involve the rigorous validation of AI-generated images within clinical environments to assess their practical utility and reliability for medical diagnostics. Given that cancer detection in medical images directly impacts human lives, it's essential to validate synthetic images thoroughly, especially when they're used to overcome the scarcity of training data. To achieve this, we will explore and apply sophisticated approaches such as t-distributed stochastic neighbor embedding (t-SNE).

Collaboration with medical professionals, particularly radiologists, will be pivotal in fine-tuning the generation processes to align closely with clinical requirements. Conducting comprehensive clinical trials will form a core part of our research to ensure our methods are practically applicable. Additionally, we will collect feedback from a wide range of potential users, including medical experts and patients, to evaluate the perceived quality and applicability of the generated images in real-life clinical settings. Using a diverse array of benchmark datasets will ensure a robust and effective training regime for our GAN models, leading to more realistic and clinically relevant mammographic images.

Exploring the application of Vision Transformers (ViT) for mammographic data analysis and image classification is another exciting avenue we plan to pursue. We will also investigate techniques that integrate multiple mammogram views, such as Left and Right CC and Left and Right MLO, for a more comprehensive analysis.

Another promising area is ensemble learning, particularly the combination of DL algorithms. This is an intriguing and relatively unexplored area within machine learning that we believe is worthwhile pursuing. We will refine various classification models before integrating them into ensemble methods to optimize our results. Additionally, leveraging the power of transfer learning by fine-tuning pre-trained models on extensive datasets is expected to yield better outcomes.

A significant advancement we anticipate is the implementation of three-dimensional reconstructions of mammograms. This approach offers a more comprehensive perspective of breast tissue, potentially revolutionizing the way mammograms are assessed and interpreted in clinical settings. It promises to enhance the diagnostic capabilities of radiologists and improve overall patient care.

## References

- [1] M. B. Buntin, M. F. Burke, M. C. Hoaglin, and D. Blumenthal, "The benefits of health information technology: a review of the recent literature shows predominantly positive results," *Health Affairs*, vol. 30, no. 3, pp. 464-471, 2011.
- [2] K. Häyrinen, K. Saranto, and P. Nykänen, "Definition, structure, content, use and impacts of electronic health records: a review of the research literature," *International Journal of Medical Informatics*, vol. 77, no. 5, pp. 291-304, 2008.
- [3] P. J. Scott, M. Rigby, E. Ammenwerth, J. Brender McNair, and A. V. Stowell, "Evaluation considerations for secondary uses of clinical data: principles for an evidence-based approach to policy and implementation of secondary analysis," *Yearbook of Medical Informatics*, vol. 26, no. 1, pp. 59-67, 2017.
- [4] S. Dang and N. Vialaneix, "Cutting Edge Bioinformatics and Biostatistics Approaches Are Bringing Precision Medicine and Nutrition to a New Era," *Lifestyle Genomics*, vol. 11, no. 2, pp. 73-76, 2018.
- [5] M. A. Kumar, R. Vimala, and K. A. Britto, "A cognitive technology based healthcare monitoring system and medical data transmission," *Measurement*, vol. 146, pp. 322-332, 2019.
- [6] American Cancer Society, "American Cancer Society," 2021. [Online]. Available: <https://www.cancer.org/>. Accessed: Nov. 15, 2021.
- [7] C. J. D'Orsi, "Breast imaging," *Radiologic Clinics*, vol. 42, no. 5, pp. 11-12, 2004.
- [8] L. Wilkinson and T. Gathani, "Understanding breast cancer as a global health concern," *The British journal of radiology*, vol. 95, no. 1130, p. 20211033, 2022.
- [9] O. Ginsburg, C. H. Yip, A. Brooks, A. Cabanes, M. Caleffi, J. A. Dunstan Yataco, *et al.*, "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, no. Suppl. 10, pp. 2379–2393, 2020.
- [10] S. A. Kurt, K. Ozgokce, M. Hattapoglu, M. Basak, and T. E. Toprak, "Prediction of molecular subtypes using superb microvascular imaging and shear wave elastography in invasive breast carcinomas," *Ultrasound Quarterly*, vol. 30, no. 1, pp. 14–21, 2023.

- [11] R. Mehrotra and K. Yadav, "Breast cancer in India: Present scenario and the challenges ahead," *World Journal of Clinical Oncology*, vol. 13, no. 3, pp. 209–218, 2022.
- [12] M. J. Ali, B. Raza, A. R. Shahid, F. Mahmood, M. A. Yousuf, A. H. Dar, and U. Iqbal, "Enhancing breast pectoral muscle segmentation performance by using skip connections in fully convolutional network," *International Journal of Imaging Systems and Technology*, vol. 30, no. 4, pp. 1108–1118, 2020.
- [13] D. C. Grossman, S. J. Curry, D. K. Owens, M. J. Barry, K. W. Davidson, A. Doubeni, *et al.*, "Screening for ovarian cancer: US preventive services task force recommendation statement," *Journal of the American Medical Association*, vol. 319, no. 6, pp. 588–594, 2018.
- [14] J. Wallyn, N. Anton, S. Akram, and T. F. Vandamme, "Biomedical imaging: Principles, technologies, clinical aspects, contrast agents, limitations, and future trends in nanomedicines," *Pharmaceutical Research*, vol. 36, pp. 1–31, 2019.
- [15] S. Iranmakani, T. Mortezaazadeh, F. Sajadian, M. F. Ghaziani, A. Ghafari, D. Khezerloo, and A. E. Musa, "A review of various modalities in breast imaging: Technical aspects and clinical outcomes," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 51, pp. 1–22, 2020.
- [16] W. E. Fathy and A. S. Ghoneim, "A deep learning approach for breast cancer mass detection," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, pp. 175–182, 2019.
- [17] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [18] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.
- [19] T. Ramesh and V. Santhi, "Exploring big data analytics in health care," *International Journal of Intelligent Networks*, vol. 1, pp. 135–140, 2020.
- [20] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature medicine*, vol. 25, no. 1, pp. 44–56, 2019.
- [21] D. C. Elton, Z. Boukouvalas, M. D. Fuge, and P. W. Chung, "Deep learning for molecular design—a review of the state of the art," *Molecular Systems Design & Engineering*, vol. 4, no. 4, pp. 828–849, 2019.

- [22] M. K. Jessen, M. F. Vallentin, M. J. Holmberg, M. Bolther, F. Hansen, and J. M. Holst, "Goal-directed haemodynamic therapy during general anaesthesia for noncardiac surgery: a systematic review and meta-analysis," *British Journal of Anaesthesia*, vol. 128, no. 3, pp. 416-433, 2022.
- [23] S. S. Han, I. Park, S. E. Chang, W. Lim, M. S. Kim, G. H. Park, *et al.*, "Augmented intelligence dermatology: deep neural networks empower medical professionals in diagnosing skin cancer and predicting treatment options for 134 skin disorders," *Journal of Investigative Dermatology*, vol. 140, no. 9, pp. 1753-1761, 2020.
- [24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [25] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097-1105, 2012.
- [26] T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, "Challenges of deep learning in medical image analysis—improving explainability and trust," *IEEE Transactions on Technology and Society*, vol. 4, no. 1, pp. 68-75, 2023.
- [27] X. Mao, W. He, K. Humphreys, M. Eriksson, N. Holowko, H. Yang, J. Tapia, P. Hall, and K. Czene, "Breast cancer incidence after a false-positive mammography result," *JAMA Oncology*, vol. 10, no. 1, pp. 63–70, 2024.
- [28] F. A. Zeiser, C. A. da Costa, T. Zonta, N. M. C. Marques, A. V. Roehle, M. Moreno, and R. da Rosa Righi, "Segmentation of masses on mammograms using data augmentation and deep learning," *Journal of Digital Imaging*, vol. 33, no. 4, pp. 858–868, 2020.
- [29] S. Guan and M. Loew, "Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks," *Journal of Medical Imaging*, vol. 6, no. 3, pp. 031411, 2019.
- [30] B. Swiderski, L. Gielata, P. Olszewski, S. Osowski, and M. Kołodziej, "Deep neural system for supporting tumor recognition of mammograms using modified GAN," *Expert Systems with Applications*, vol. 164, p. 113968, 2021.
- [31] N. F. Razali, I. S. Isa, S. N. Sulaiman, N. K. A. Karim, and M. K. Osman, "High-level features in deeper deep learning layers for breast cancer classification," in *Proc. 2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCCE)*, 2021, pp. 170–175.

- [32] D. Ueda, A. Yamamoto, N. Onoda, T. Takashima, S. Noda, S. Kashiwagi *et al.*, "Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets," *PLoS One*, vol. 17, no. 3, p. e0265751, 2022.
- [33] Y. J. Kim and K. G. Kim, "Detection and weak segmentation of masses in gray-scale breast mammogram images using deep learning," *Yonsei Medical Journal*, vol. 63, no. Suppl, p. S63-S73, 2022.
- [34] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO based breast masses detection and classification in full-field digital mammograms," *Computer methods and programs in biomedicine*, vol. 200, p. 105823, 2021.
- [35] A. Sabani, A. Landsmann, P. Hejduk, C. Schmidt, M. Marcon, K. Borkowski, C. Rossi, A. Ciritsis, and A. Boss, "BI-RADS-based classification of mammographic soft tissue opacities using a deep convolutional neural network," *Diagnostics*, vol. 12, no. 7, p. 1564, 2022.
- [36] E. Wu, K. Wu, and W. Lotter, "Synthesizing lesions using contextual GANs improves breast cancer classification on mammograms," *arXiv preprint arXiv:2006.00086*, 2020.
- [37] S. D. Desai, S. Giraddi, N. Verma, P. Gupta, and S. Ramya, "Breast cancer detection using gan for limited labeled dataset," in *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2020, pp. 34-39.
- [38] B. Alyafi, O. Diaz, and R. Marti, "DCGANs for realistic breast mass augmentation in x-ray mammography," in *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314, SPIE, pp. 473–480, 2020.
- [39] T. Shen, K. Hao, C. Gou, and F.-Y. Wang, "Mass image synthesis in mammogram with contextual information based on GANs," *Computer Methods and Programs in Biomedicine*, vol. 202, p. 106019, 2021.
- [40] J. Lee and R. M. Nishikawa, "Identifying women with mammographically-occult breast cancer leveraging GAN-simulated mammograms," *IEEE transactions on medical imaging*, vol. 41, no. 1, pp. 225-236, 2021.
- [41] S. Park, K. H. Lee, B. Ko, and N. Kim, "Unsupervised anomaly detection with generative adversarial networks in mammography," *Scientific Reports*, vol. 13, no. 1, p. 2925, 2023.

- [42] P. M. Shah, H. Ullah, R. Ullah, D. Shah, Y. Wang, S. ul Islam, A. Gani, and J. P. C. Rodrigues, "DC-GAN-based synthetic X-ray images augmentation for increasing the performance of EfficientNet for COVID-19 detection," *Expert Systems*, vol. 39, no. 3, p. e12823, 2022.
- [43] S. Yang, K.-D. Kim, E. Ariji, N. Takata, and Y. Kise, "Evaluating the performance of generative adversarial network-synthesized periapical images in classifying C-shaped root canals," *Scientific Reports*, vol. 13, no. 1, p. 18038, 2023.
- [44] A. Figueira and B. Vaz, "Survey on Synthetic Data Generation, Evaluation Methods and GANs," *Mathematics*, vol. 10, no. 15, p. 2733, 2022.
- [45] S. Meister, N. Möller, J. Stüve, and R. M. Groves, "Synthetic image data augmentation for fibre layup inspection processes: Techniques to enhance the data set," *Journal of Intelligent Manufacturing*, vol. 32, no. 6, pp. 1767-1789, 2021.
- [46] D. Korkinof, H. Harvey, A. Heindl, E. Karpati, G. Williams, T. Rijken, P. Kecskemethy, and B. Glocker, "Perceived realism of high-resolution generative adversarial network-derived synthetic mammograms," *Radiology: Artificial Intelligence*, vol. 3, no. 2, p. e190181, 2020.
- [47] M. Sanchez, M. Brieu, M. Cosson, and P. Nielsen, "Patient-specific biomechanical modeling for applications in breast cancer diagnosis and treatment," in *Biomechanics of the Female Reproductive System: Breast and Pelvic Organs*, M. Brieu, Ed. pp. 333–356, 2023.
- [48] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. H. Mak, and H. J. Aerts, "Deep learning predicts lung cancer treatment response from serial medical imaging," *Clinical Cancer Research*, vol. 25, no. 11, pp. 3266-3275, 2019.
- [49] L. Gonzalez-Abril, C. Angulo, J. A. Ortega, and J.-L. Lopez-Guerra, "Statistical validation of synthetic data for lung cancer patients generated by using generative adversarial networks," *Electronics*, vol. 11, no. 20, p. 3277, 2022.
- [50] K. H. Cha, N. A. Petrick, A. Pezeshk, C. G. Graff, D. Sharma, A. Badal, and B. Sahiner, "Evaluation of data augmentation via synthetic images for improved breast mass detection on mammograms using deep learning," *Journal of Medical Imaging*, vol. 7, no. 1, pp. 012703-012703, 2020.

- 
- [51] I. Furtney, R. Bradley, and M. R. Kabuka, "Patient Graph Deep Learning to Predict Breast Cancer Molecular Subtype," *IEEE/ACM transactions on computational biology and bioinformatics*, 2023.
  - [52] E. Shivhare and V. Saxena, "Optimized generative adversarial network based breast cancer diagnosis with wavelet and texture features," *Multimedia Systems*, vol. 28, no. 5, pp. 1639-1655, 2022.
  - [53] F. Jacobs, G. Saltalamacchia, C. Benvenuti, S. D'Amico, M. Gaudio, C. Miggiano, A. Zambelli, "Opportunities and challenges of synthetic data generation in oncology," *JCO Clinical Cancer Informatics*, vol. 7, p. e2300045, 2023.
  - [54] B. Savelli, A. Bria, M. Molinara, C. Marrocco, and F. Tortorella, "A multi-context cnn ensemble for small lesion detection," *Artificial intelligence in medicine*, vol. 103, p. 101749, 2020.
  - [55] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, pp. 6105-6114.
  - [56] L. Li, M. Xu, X. Wang, L. Jiang, and H. Liu, "Attention based glaucoma detection: A large-scale database and cnn model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10571-10580.
  - [57] A. Khamparia, S. Bharati, P. Podder, D. Gupta, A. Khanna, T. K. Phung, and D. N. H. Thanh, "Diagnosis of breast cancer based on modern mammography using hybrid transfer learning," *Multidimensional systems and signal processing*, vol. 32, no. 2, pp. 747-765, 2021.
  - [58] A. Pillai, A. Nizam, M. Joshee, A. Pinto, and S. Chavan, "Breast cancer detection in mammograms using deep learning," in *Applied Information Processing Systems: Proceedings of ICCET 2021*, pp. 121-127. Springer Singapore, 2022.
  - [59] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data," *MethodsX*, vol. 7, p. 100864, 2020.
  - [60] S. Kuo and O. Honda, "Mammographic mass detection based on data separated ensemble convolution neural network," in *2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2021, pp. 432-437.



- [61] D. Vijayan and R. Lavanya, "Ensemble of density-specific experts for mass characterization in mammograms," *J Signal, Image Video Processing*, vol. 15, no. 5, pp. 1011-1019, 2021.
- [62] S. Yu, L. Liu, Z. Wang, G. Dai, and Y. Xie, "Transferring deep neural networks for the differentiation of mammographic breast lesions," *Science China Technological Sciences*, vol. 62, pp. 441-447, 2019.
- [63] S. Boumaraf, X. Liu, C. Ferkous, and X. Ma, "A new computer-aided diagnosis system with modified genetic feature selection for BI-RADS classification of breast masses in mammograms," *BioMed Research International*, vol. 2020, Article ID 7695207, 2020.
- [64] A. H. Farhan and M. Y. Kamil, "Texture analysis of breast cancer via LBP, HOG, and GLCM techniques," in *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 7, p. 072098, 2020.
- [65] R. Song, T. Li, and Y. Wang, "Mammographic classification based on XGBoost and DCNN with multi features," *IEEE Access*, vol. 8, pp. 75011-75021, 2020.
- [66] R. Arora, P. K. Rai, and B. Raman, "Deep feature-based automatic classification of mammograms," *Medical & biological engineering & computing*, vol. 58, pp. 1199-1211, 2020.
- [67] H. Zhang, R. Wu, T. Yuan, Z. Jiang, S. Huang, J. Wu, J. Hua, Z. Niu, and D. Ji, "DE-Ada\*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions," *Information Sciences*, vol. 539, pp. 461-486, 2020.
- [68] R. S. Patil and N. Biradar, "Automated mammogram breast cancer detection using the optimized combination of convolutional and recurrent neural network," *Evolutionary intelligence*, vol. 14, pp. 1459-1474, 2021.
- [69] H. Li, R. Mukundan, and S. Boyd, "Novel texture feature descriptors based on multi-fractal analysis and lbp for classifying breast density in mammograms," *Journal of Imaging*, vol. 7, no. 10, p. 205, 2021.
- [70] M. U. Farooq, Z. Ullah, and J. Gwak, "Ensemble CNNs for Breast Tumor Classification," *arXiv preprint arXiv:2304.13727*, 2023.
- [71] A. Mohamed, E. Amer, N. Eldin, M. Hossam, N. Elmasry, and G. T. Adnan, "The impact of data processing and ensemble on breast cancer detection using deep learning," *Journal of Computing and Communication*, vol. 1, no. 1, pp. 27-37, 2022.

- [72] N. F. Abubacker, I. A. T. Hashem, and L. K. Hui, "Mammographic classification using stacked ensemble learning with bagging and boosting techniques," *Journal of Medical and Biological Engineering*, vol. 40, pp. 908-916, 2020.
- [73] F. Z. Nakach, A. Idri, and H. Zerouaoui, "Deep hybrid bagging ensembles for classifying histopathological breast cancer images," in *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART)*, 2023, pp. 289-300.
- [74] A. A. Hekal, H. E.-D. Moustafa, and A. Elnakib, "Ensemble deep learning system for early breast cancer detection," *Evolutionary Intelligence*, vol. 16, no. 3, pp. 1045-1054, 2023.
- [75] C. Scapicchio, F. Lizzi, and M. Fantacci, "Explainability of a CNN for breast density assessment," *Il Nuovo Cimento*, vol. 44, no. 4-5, pp. 1-4, 2021.
- [76] E. Pintelas, I. E. Livieris, and P. Pintelas, "A grey-box ensemble model exploiting black-box accuracy and white-box intrinsic interpretability," *Journal of Algorithms*, vol. 13, no. 1, p. 17, 2020.
- [77] R. M. Mann, A. Athanasiou, P. A. T. Baltzer, J. Camps-Herrero, P. Clauser, E. M. Fallenber, *et al.*, "Breast cancer screening in women with extremely dense breasts recommendations of the European Society of Breast Imaging (EUSOBI)," *European radiology*, vol. 32, no. 6, pp. 4036-4045, 2022.
- [78] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021.
- [79] F. De Bueger, M. Decroës, and B. Macq, "Deep learning in mammography: Reducing annotation effort for breast mass segmentation," Master's thesis, Ecole Polytechnique de Louvain, Université Catholique de Louvain, 2022. [Online]. Available: <https://dial.uclouvain.be/memoire/ucl/en/object/thesis%3A35700..>
- [80] S. International, "Importance of breast cancer awareness in Pakistan," vol. 2023, ed., 2022.
- [81] "Mammography," National Institute of Biomedical Imaging and Bioengineering. [Online]. Available: <https://www.nibib.nih.gov/science-education/science-topics/mammography>. [Accessed: 12-Mar-2023].

- [82] N. Sharma, A. Y. Ng, J. J. James, G. Khara, É. Ambrózay, C. C. Austin, G. Forrai, G. Fox, B. Glocker, A. Heindl, *et al.*, "Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms," *BMC cancer*, vol. 23, no. 1, pp. 1-13, 2023.
- [83] L. Abdelrahman, M. Al Ghamdi, F. Collado-Mesa, and M. Abdel-Mottaleb, "Convolutional neural networks for breast cancer detection in mammography: A survey," *Computers in biology and medicine*, vol. 131, p. 104248, 2021.
- [84] I. Zyout, "Evaluation of deep convolutional neural networks for detecting nonpalpable breast abnormalities in mammography," in 2023 Advances in Science and Engineering Technology International Conferences (ASET), 2023, pp. 1-6.
- [85] P. Oza, P. Sharma, S. Patel, F. Adedoyin, and A. Bruno, "Image augmentation techniques for mammogram analysis," *Journal of Imaging*, vol. 8, no. 5, p. 141, 2022.
- [86] F. Garcea, A. Serra, F. Lamberti, and L. Morra, "Data augmentation for medical imaging: A systematic literature review," *Computers in Biology and Medicine*, vol. 152, p. 106391, 2023.
- [87] D. Cömert, C. H. van Gils, W. B. Veldhuis, and R. M. Mann, "Challenges and changes of the breast cancer screening paradigm," *Journal of Magnetic Resonance Imaging*, vol. 57, no. 3, pp. 706-726, 2023.
- [88] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," *Journal of Imaging*, vol. 9, no. 4, p. 81, 2023.
- [89] University of South Florida, "Digital Database for Screening Mammography," [Online]. Available: <http://www.eng.usf.edu/cvprg/mammography/database.html>. [Accessed: 12-Mar-2023]
- [90] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [91] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, *et al.*, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *The Breast*, vol. 66, pp. 15-23, 2022.

- [92] "Breast cancer," World Health Organization. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. [Accessed: 12-Mar-2023].
- [93] N. Howlader, C. Desantis, H. Rungay, A. Mafra, D. Singh, M. Laversanne, *et al.*, "Overview of breast cancer collaborative stage data items—their definitions, quality, usage, and clinical implications: A review of SEER data for 2004-2010," *Cancer*, vol. 120, pp. 3771-3780, 2014.
- [94] M. Løberg, M. L. Lousdal, M. Bretthauer, and M. Kalager, "Benefits and harms of mammography screening," *Breast Cancer Research*, vol. 17, no. 1, pp. 1-12, 2015.
- [95] P. Oza, Y. Shah, and M. Vegda, "A comprehensive study of mammogram classification techniques," in *Tracking and Preventing Diseases with Artificial Intelligence*, Cham: Springer, 2022, pp. 217–238.
- [96] M. Elbatel, "Mammograms classification: A review," *arXiv preprint arXiv:2203.03618*, 2022.
- [97] A. Pillai, A. Nizam, M. Joshee, A. Pinto, and S. Chavan, "Breast cancer detection in mammograms using deep learning," in *Applied Information Processing Systems: Proceedings of ICCET 2021*, 2022, pp. 121-127.
- [98] M. Dib, N. J. Oliveira, A. E. Marques, M. C. Oliveira, J. V. Fernandes, and B. M. Ribeiro, "Single and ensemble classifiers for defect prediction in sheet metal forming under variability," *Neural Computing and Applications*, vol. 32, pp. 12335-12349, 2020.
- [99] S. Saikia, V. P. V, B. G. Prajapati, J. Prajapati, A. Parihar, and R. Malviya, "Computational tools for drug discovery of anticancer therapy," in *Targeted Cancer Therapy in Biomedical Engineering*, Springer, 2023, pp. 887-904..
- [100] F. Yan, H. Huang, W. Pedrycz, and K. Hirota, "Automated breast cancer detection in mammography using ensemble classifier and feature weighting algorithms," *Expert Systems with Applications*, vol. 227, p. 120282, 2023/10/01/ 2023.
- [101] N. M. Hassan, S. Hamad, and K. Mahar, "Mammogram breast cancer CAD systems for mass detection and classification: a review," *Multimedia Tools and Applications*, vol. 81, no. 14, pp. 20043-20075, 2022.
- [102] F. Chris Carr, George Partridge, inversion, Jayashree Kalpathy-Cramer, John Mongan, Katherine Andriole, Lavender, Maryam Vazirabad, Michelle Riopel,

- Robyn Ball, Sohier Dane, Yan Chen. *RSNA Screening Mammography Breast Cancer Detection*. [Online]. Available: <https://kaggle.com/competitions/rsna-breast-cancer-detection>
- [103] X. Wang, I. Ahmad, D. Javeed, S. A. Zaidi, F. M. Alotaibi, M. E. Ghoneim, Y. I. Daradkeh, J. Asghar, and E. T. Eldin, "Intelligent Hybrid Deep Learning Model for Breast Cancer Detection," *Electronics*, vol. 11, no. 17, p. 2767, 2022.
- [104] A. Dibden, J. Offman, S. W. Duffy, and R. Gabe, "Worldwide review and meta-analysis of cohort studies measuring the effect of mammography screening programmes on incidence-based breast cancer mortality," *Cancers*, vol. 12, no. 4, p. 976, 2020.
- [105] F. F. Ting, Y. J. Tan, and K. S. Sim, "Convolutional neural network improvement for breast cancer classification," *Expert Systems with Applications*, vol. 120, pp. 103-115, 2019.
- [106] L. Bouzar-Benlabiod, K. Harrar, L. Yamoun, M. Y. Khodja, and M. A. Akhloufi, "A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification," *Computers in Biology and Medicine*, vol. 163, p. 107133, 2023.
- [107] M. A. Saleem, N. Senan, F. Wahid, M. Aamir, A. Samad, and M. Khan, "Comparative analysis of recent architecture of Convolutional Neural Network," *Mathematical Problems in Engineering*, vol. 2022, 2022.
- [108] WHO, "Fact Sheet," 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets>. [Accessed: Oct. 2023].
- [109] A. Gastounioti, S. Desai, V. S. Ahluwalia, E. F. Conant, and D. Kontos, "Artificial intelligence in mammographic phenotyping of breast cancer risk: A narrative review," *Breast Cancer Research*, vol. 24, no. 1, pp. 1-12, 2022.
- [110] A. S. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, "Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer," *Investigative radiology*, vol. 52, no. 7, pp. 434-440, 2017.
- [111] M. A. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, 2022.

- [112] M. Heath, K. Bowyer, D. Kopans, W. P. Kegelmeyer, R. Moore, and K. Chang, "Current status of the digital database for screening mammography," in *Digital Mammography: Nijmegen, 1998*: Springer, 1998, pp. 457-460.
- [113] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [114] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.
- [115] Y. Bi, J. Guan, and D. Bell, "The combination of multiple classifiers using an evidential reasoning approach," *Artificial Intelligence*, vol. 172, no. 15, pp. 1731-1751, 2008.
- [116] Y. Bi, "The impact of diversity on the accuracy of evidential classifier ensembles," *International Journal of Approximate Reasoning*, vol. 53, no. 4, pp. 584-607, 2012.
- [117] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921-2929.
- [118] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.
- [119] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097-2106.
- [120] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60-88, 2017.
- [121] J. E. Thatcher, J. J. Squiers, S. C. Kanick, D. R. King, Y. Lu, Y. Wang, R. Mohan, E. W. Sellke, and J. M. DiMaio, "Imaging techniques for clinical burn

- assessment with a focus on multispectral imaging," *Advances in wound care*, vol. 5, no. 8, pp. 360-378, 2016.
- [122] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221-248, 2017.
- [123] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.
- [124] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618-626.
- [125] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," in *2018 IEEE winter conference on applications of computer vision (WACV)*, 2018, pp. 839-847.